CENTER FOR COGNITIVE SCIENCE

The Center for Cognitive Science at MIT was established March 1, 1979 in order to facilitate interdisciplinary studies of mental representation, cognitive processing, and the biological representation of cognition. Members of the Center include faculty and staff from a variety of departments, divisions, and laboratories within the Institute, as well as a number of Visiting Fellows.

For additional copies of this document or further information about the Center, please contact:

Director,
The Center for Cognitive Science
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-98-

0732

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1990	Final

4. TITLE AND SUBTITLE
Overregularization

5. FUNDING NUMBERS

6. AUTHORS
Gary F. Marcus, Michael Ullman, Steven Pinker, Michelle Hollander, T. John Rosen, Fei Xu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Massachusetts Institute of Technology

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NI
4040 Fairfax Dr, Suite 500
Arlington, VA 22203-1613

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
See attachment

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

DTIC QUALITY INSPECTED 3

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239.18
Designed using WordPerfect 6.1, AFOSR/XPP, Oct 96

Occasional Paper #41

Overregularization

Gary F. Marcus
Michael Ullman
Steven Pinker
Michelle Hollander
T. John Rosen
Fei Xu

November 1990

Gary Marcus, Michael Ullman, Steven Pinker, Michelle Hollander, and T. John Rosen are in the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology. Fei Xu is in the Department of Psychology, Smith College.

We thank John J. Kim, Sandeep Prasada, and Alan Prince for many helpful discussions and suggestions, and Jill de Villiers, and Beth Levin, Michael Maratsos, Dan Slobin, and Joseph Stemberger, for sharing data and manuscripts. Parts of this paper were presented at the 1989 and 1990 Boston University Conference on Language Development. Supported by NIH Grant HD 18381 to the third author and by a grant from the Alfred P. Sloan Foundation to the MIT Center for Cognitive Science. The first and second authors are supported by NDSE Graduate Fellowships. The last author was supported by a Dana Foundation Award. Requests for reprints should be sent to the first author at E10-109, MIT, Cambridge MA 02139.

Table of Contents

Abstract

Children's overregularization errors such as *comed* bear on three issues: "U"-shaped development where children get worse over time because of an interaction between memory and rule-governed processes; the unlearning of grammatical errors in the absence of parental negative feedback; and whether cognitive processes are computed by rules or by parallel distributed processing (connectionist) networks. We remedy the lack of quantitative data on overregularization by exhaustively analyzing the 11,500 irregular past tense utterances in the transcribed spontaneous speech of 69 children, and by reviewing the naturalistic and experimental literature. We found: (1) overregularization errors are relatively rare (median 2.5% of irregular past tense forms), suggesting that there is no qualitative defect in children's grammars that must be unlearned. (2) Overregularization occurs at a roughly constant low rate from the late two's into the school-age years, affecting most irregular verbs. (3) Though there is no stage where overregularization errors predominate, one other aspect of U-shaped development was confirmed: an extended period of correct performance before the first overregularization. (4) No support was found for Rumelhart & McClelland's (1986) hypothesis that overregularization is caused by increases in the number or proportion of regular verbs in the input to the past tense system (either parents' tokens, children's tokens, or children's types). Thus the traditional account in which a memory system operates before a rule system cannot be replaced by a connectionist alternative in which a single network displays rotelike or rulelike behavior in response to changes in input statistics. (5) The onset of overregularization is best predicted by the onset of *obligatoriness*: the errors appear when children stop leaving verbs in past tense contexts unmarked (e.g., *Yesterday I come*). (6) The more often a parent uses an irregular past tense form of a verb, the less often the child overregularizes it. (7) Verbs are protected from overregularization by neighborhoods of similar-sounding irregulars, but are not attracted to overregularization by neighborhoods of similar-sounding regulars. This suggests that the associative properties of connectionist networks may help explain performance with irregulars (via the memory system in which they are stored) but not with regulars. A simple hypothesis explains these phenomena. Children, like adults, obligatorily mark tense, using one of two mechanisms: memory for irregulars, and an affixation rule that can generate a regular past tense form for any verb. Retrieval of an irregular blocks the rule, but children's memory traces for irregulars are not strong enough to guarantee perfect retrieval. When retrieval fails, the rule is applied, and overregularization results.

Overregularization

1. Introduction

Overregularizations like *comed* and *foots* are among the most conspicuous grammatical errors in child language, and they have been commented upon for as long as language development has been studied (Chamberlain, 1906; Bateman, 1916; Smith, 1933; Carlton, 1947; Carroll, 1961; Brown & Bellugi, 1964; Ervin & Miller, 1963; Ervin, 1964; Guillaume, 1927; Menyuk, 1963; Miller & Ervin, 1964; Cazden, 1968; Slobin, 1971; Brown, 1973; Kuczaj, 1977, 1981; Slobin, 1978; Bybee & Slobin, 1982a; see Edwards, 1970). These errors are made possible by the fact that English has two ways of creating past tense forms: Most verbs add the suffix *-ed* to their stems, but about 180 exceptional or "irregular" verbs form their past tenses in idiosyncratic ways such as a vowel change (*come-came*), replacement of a final consonant or rhyme (*make-made, teach/taught*), substitution of another form (*go-went*), or no change at all (*cut-cut*). Overregularization errors consist of applying the regular pattern to an irregular stem.

Since children do not hear these forms from their parents, the errors reveal the operation of a creative process, presumably corresponding to a mental operation implementing the *-ed*-suffixation rule posited by grammarians. As such the errors are often offered as the quintessential demonstration of the creative essence of human language (Chomsky, 1959), and of the necessity of explaining cognitive processes by rules and representations rather than by rote and reinforcement (e.g., Brown & Bellugi, 1964; Lenneberg, 1964; McNeill, 1966; Slobin, 1971). The productive mental process responsible for overregularization reveals itself in other ways. Children frequently inflect their own invented verbs such as *speeched* (Chamberlain, 1906), *by-ed* (= "went by"; Miller & Ervin, 1964); *eat lunched* (Kuczaj, 1977), *broomed* (Clark, 1982), and *grained* (Pinker, Lebeaux, & Frost, 1987). In many experiments, beginning with Berko's (1958) classic study, children are given a nonce verb and are then asked to use it in a past tense context, such as "Here is a man who likes to rick. He did the same thing yesterday. Yesterday he ____." The children readily produce appropriate forms such as *ricked*, and when provided with an existing irregular stem, frequently overregularize it (Berko, 1958; Miller & Ervin, 1964; Anisfeld & Tucker, 1967; Bryant & Anisfeld, 1969; Kuczaj, 1978; Derwing & Baker, 1979; Pinker, Lebeaux, & Frost, 1987; Marchman, 1988; Cox, 1989; Kim, Marcus, Hollander, and Pinker, in preparation.) In other experiments they have been found to judge overregularizations (Kuczaj, 1978) or nonce forms resembling regularly inflected forms (Anisfeld, Barlow, & Frail, 1968; Anisfeld & Gordon, 1968) as acceptable.

2. Issues Raised by Overregularization

Overregularization figures prominently in three areas of current research: the interaction between rote and rules in development, especially "U"-shaped development where children get worse over time; the unlearning of grammatical errors in the absence of negative feedback from parents; and whether cognitive processes are computed by rules or by parallel distributed processing (connectionist) networks. We will review them in detail, to highlight the empirical gaps that prevent the issues from being resolved given currently available data.

2.1. U-shaped Development

An interesting feature of overregularization, first noted by Ervin & Miller (1963; see also Miller & Ervin, 1964; Cazden, 1968; and Pinker & Prince, 1988), is that it follows a period in which children produce the irregular past tense forms correctly; overregularization represents a decline in performance as the child gets older, resulting in a "U"-shaped curve if the proportion of irregular past tense forms that are correct is plotted against age. This nonmonotonicity seems to reveal a reorganization of the child's linguistic system, reflecting a tendency to ferret out generalizations and to prefer them to exceptional forms, a tendency that is sometimes considered paradigmatic of language development (Slobin, 1973; Bowerman, 1982) and cognitive development (Strauss, 1982) in general.

MacWhinney (1978) and Pinker (1984) explained this U-shaped sequence by positing a dissociation between two psychological processes: rote memory, which simply records irregulars from parental speech, and rule deployment, which can only operate once the rule itself has been abstracted from a set of regularly inflected forms. The rote process operates on a verb-by-verb basis and allows the child to use a given irregular form correctly at the outset of language development, but rule deployment must await the abstraction of the rule itself from a set of regular forms accumulated over time from the parental input. As we shall see, there are alternatives to this account.

2.2. Negative Evidence and Recovery from Errors

A second research problem involves the unlearning of the overregularization errors some time before adulthood (the right-hand arm of the "U"). This has been a focus of research in learnability-theoretic approaches to language development, which aim to identify the learning mechanisms by which the child successfully attains full knowledge of the adult language (Pinker, 1979; Wexler & Culicover, 1980; Osherson, Stob, and Weinstein, 1985). A significant problem in explaining acquisition is that children do not receive significant "negative evidence": feedback from parents indicating, for any string of words they may utter, whether it is a grammatical sentence. Children are not corrected or misunderstood more often when they speak ungrammatically (Brown & Hanlon, 1970), and it is doubtful whether other forms of parental reactions, such as differential likelihood of repetitions, expansions, or topic changes, provide useful information to the child for the vast majority of grammatical violations (see Bowerman, 1987; Gordon, 1990; Pinker, 1989; Grimshaw & Pinker, 1989). A lack of negative evidence means that if the child ever develops a linguistic system that generates a superset of the target language, the parental input cannot tell the child anything is wrong (Gold, 1967). To explain how the adult grammar is attained, then, one must explain either how the children avoid generating supersets, or, if they do, how they expunge their errors.

Overregularization errors in particular pose this problem. Kuczaj (1977, p. 599) noted that in his investigation the children (especially his son Abe who was the main subject) were not corrected for overregularization errors. The following typical chunk of dialogue, which we have found in transcripts of conversations between Kuczaj and Abe (MacWhinney & Snow, 1985), illustrates the claim:

Father: Where is that big piece of paper I gave you yesterday?
Abe: Remember? I writed on it.
Father: Oh that's right don't you have any paper down here buddy?

Moreover it seems unlikely that children attend to corrections, requests for clarification, recastings, and

so on, when they do occur; the following dialogue, from Cazden (1972), is typical:

Child: My teacher holded the baby rabbits and we patted them.
Adult: Did you say your teacher held the baby rabbits?
Child: Yes.
Adult: What did you say she did?
Child: She holded the baby rabbits and we patted them.
Adult: Did you say she held them tightly?
Child: No, she holded them loosely.

More precisely, Morgan and Travis (1989) analyzed parental responses to overregularization errors in the speech of the children known as Adam, Eve, and Sarah (Brown 1973). They focused specifically on overregularization errors, together with errors in *wh*-questions, and tabulated the proportions of grammatical and ungrammatical sentences by each child that were followed by parental expansions, exact imitations, partial imitations, clarification questions, confirmation questions, conversational "move-ons," and no response. No consistent contingency was found: for Adam, expansions and clarification questions were more likely to follow his ungrammatical sentences; for Eve it was expansions and partial imitations that occurred more frequently following her ungrammatical sentences; and for Sarah all five response categories of parental response showed the *opposite* pattern, occurring more frequently after well-formed utterances. Unless a child can figure out the kind of parent he or she has, such feedback is useless. Moreover, Morgan and Travis showed that as children get older, even this inconsistent feedback signal becomes more diluted. Since the errors continue after the parent has stopped supplying potential feedback, learning must depend on some other information source, presumably an endogenous one.

To understand how children unlearn overregularization errors, one must first examine how adults avoid them, in spite of possessing a productive regular inflection process. For the great majority of irregular verbs adults not only use the irregular past but reject the regularized version as ungrammatical (e.g., **comed, *breaked;* see Pinker & Prince, 1988). Therefore their productive regular process must be blocked from applying to any verb that has an irregular past tense form listed in the lexicon. (The regular version of verbs that do have two past tense forms, as in *dived* and *dove,* would be recorded in the lexicon from the input just as if it were an irregular; see Pinker, 1984; Ullman and Pinker, 1990.) This is sometimes called the *Blocking* or *Unique Entry* principle (Aronoff, 1976; Kiparsky, 1982; Pinker, 1984), and is related to the Principle of Contrast (Clark, 1987), which requires all contrasting surface forms in a language to differ in meaning, and the Uniqueness Principle (Wexler & Culicover, 1980; Pinker, 1984), which requires the child to generate a single surface form for each underlying form unless more than one form is present in the input. The simplest hypothesis is that the blocking principle prevents the child from entertaining a grammar that generates a superset of the adult inflectional system: Once the child is overregularizing, each time an irregular past tense form is heard in parental speech, the child can record it in the lexicon and the regular rule is thereafter blocked from applying to it. This would explain recovery from any overregularization, though more must be said to explain why for some verbs the correct irregular is produced before the first overregularization. The coining of the regular rule provides a past tense competitor for the listed irregular, and according to the Uniqueness or Contrast principle, one or the other must be eliminated; perhaps the listed form is eliminated as a consequence of the new rule (a form of "inflectional imperialism; Slobin, 1973), and must be relearned, whereupon it successfully blocks the regular rule.

However, this hypothesis, whereby the child always avoids generating a superset of the adult

language, has two problems, one theoretical, the other empirical. It is not enough for the Uniqueness and Contrast principles merely to say that competing surface forms cannot coexist; to account for the adult state one must add that it is the form attested in the input that always wins out over the one generated by a rule (Pinker, 1984; this is the difference between the Blocking principle, which specifically gives preference to listed forms, and the more general Uniqueness and Contrast principles). A child who respects Blocking should never allow the regular rule to expunge an irregular form; the irregular would win the Uniqueness competition from the start, and no U-shaped sequence should be seen. It is fortunate, then, that the facts do not support the avoid-supersets hypothesis and the theoretical problem with the explanation is moot. Although the picture of U-shaped development in which the child actually *loses* early irregular forms when overregularization begins is ubiquitous, especially in textbooks (e.g., Reich, 1986, p. 148), it has long been known to be false. When overregularizations occur, they coexist with, rather than replace, the early irregulars (Ervin and Miller, 1963; Cazden, 1968; Kuczaj, 1977, 1981). Therefore the problem is explaining how children recover from supersets, not how they avoid them.

Pinker (1984) discusses two solutions. One, proposed by Kuczaj (1977, 1981), is that children may fail to realize that a given irregular form corresponds to the past tense version of some stem. Rather, they may treat the irregular past as an independent verb, and Blocking would not apply. Errors would cease when the two verbs were united, presumably when the child noticed that they were semantically identical except for pastness and (in most cases) were phonologically similar as well. Two phenomena support this hypothesis. First, Kuczaj (1981) and Pinker & Prince (1988) noted that children productively inflect irregular stems: *ate* and *ating* coexist with *eated* and *eating*, as if they were two verbs. Second, Bybee and Slobin (1982a) noted that irregular verbs that end in a vowel that changes from stem to past (*fly/flew, see/saw, know/knew, blow/blew,* etc.) are particularly prone to being overregularized. If phonological overlap between stem and past is a critical cue for two forms to be lumped together as versions of the same verb, the meager common portion among verbs in this class (e.g., initial *s* for *see* and *saw*) would make it harder for the child to recognize that they were forms of a single verb.

Unfortunately, there are problems for the hypothesis that a paradigm unification failure is a major cause of overregularization. Errors like *ating* and *ated* are uncommon, even in past contexts, and are found late in development when they do occur. Furthermore the cues that would tell the child that the two forms were versions of the same verb are present throughout development, leaving it a puzzle that the child takes so long to notice them. Moreover, the linguistically valid cues for common membership militate strongly against the child's ever considering the past stem to be an independent verb. Kuczaj (1981) notes that the child correctly treats the past stem as indicating pastness: *was wenting* occurs; *is wenting* does not. But according to crosslinguistic research in lexical semantics, tense is an extremely unnatural, perhaps nonexistent, semantic component for an independent verb (Bybee, 1985; Talmy, 1985; Pinker, 1989): Languages do not like independent verbs that mean "do X" and "do X in the past." If children's hypotheses mesh with what is linguistically possible, they should not posit such verbs.

The second solution is simpler: children may always possess the correct irregular and represent it as the past of the corresponding stem, but either the memory entry for the irregular, the link to the stem, or both may not be accessible 100% of the time. Whenever the past form is not retrieved, it will not block regularization, and an overregularization will be the output (assuming that the child recognizes that

tense marking of some kind or another is obligatory in the language).[1] Pinker (1984) notes that the retrieval-failure hypothesis is consistent with the fact that adults occasionally make overregularization errors in their spontaneous speech (Stemberger, 1982). Retrieval failure is the only plausible explanation for adults, and it would apply even more readily to children, who have heard each irregular fewer times than adults. Furthermore the hypothesis is consistent with the fact that irregulars tend to be high in frequency in English, and that lower frequency irregular verbs in earlier stages of the language (e.g., *geld/gelt, cleave/clove, abide/abode*) were likely to become regular over time (Bybee, 1985). Low frequency past tense forms are always in danger of not being uniformly memorized in some generation; if so, the verbs, if they remain in the language at all, will become regular. Verbs that survive as irregulars are thus more likely to be high in frequency.

Indeed the retrieval failure hypothesis is just a combination of the logic of irregularity with the fact, known since Ebbinghaus, that human memory retrieval is probabilistic, with a higher probability of retrieval for items that have been presented to the learner more often. What is the past tense form of the verb *to shend,* meaning "to shame"? If you answered *shended* then you have overregularized; the correct form is *shent* (Bybee and Slobin, 1982b). This is not surprising; you have heard *shent* zero times. Now, if in two years you were asked the question again and overregularized once more it would still not be surprising, because you would have heard it only once, which may not enough times to consolidate it in memory. Similarly, many adults may "overregularize" *smote, slew, begot, bade, hove,* and other irregular verbs they have not recently encountered (Ullman & Pinker, 1990). A child who has heard a more common irregular verb only a few times in the past tense would be in the same situation, perhaps more often, holding number of exposures constant, if children's memory retrieval is noisier than adults'.

2.3. Rule-Based versus Connectionist Cognitive Architectures

Currently overregularization is an important topic in the very foundations of cognitive science. Virtually all accounts of the phenomenon have assumed that it can only be modeled in principle by some explicit representation of a rule. Rumelhart and McClelland (1986) showed this to be false. They devised a computer simulation of an associative network that acquired hundreds of regular and irregular verbs and generalized properly to dozens of new verbs that it was not trained on. More strikingly, it appeared to go through a U-shaped developmental sequence of correct followed by overregularized irregular verbs, and seemed to manifest several other effects previously known to characterize children's behavior. But the model has no explicit representation of words, rules, or a distinction between regular and irregular systems. Rather, it is a simple network with two layers of nodes -- a set of input units that are turned on in patterns that correspond to the sound of the verb stem, and a set of output units that are turned on in patterns that correspond to the sound of the verb's past tense form -- and weighted connections between every input unit and every output unit. Each unit corresponds to a sequence of phonological features, such as a high vowel between two stop consonants, or a back vowel followed by a nasal consonant at the end of a word; the word itself is represented by the set of feature sequences it contains. When a set of input nodes is activated, each node sends its activation level, multiplied by the link weight, to the output

[1]Overregularizations consisting of an affixed past stem, like *broked* or *wented,* would result from a slightly different kind of retrieval failure: rather than turning up nothing, the accessing process would retrieve the past tense form, but without its "past tense" feature. At that moment it would be indistinguishable from a stem and hence would be regularized.

nodes it is connected to. Each output node sums its weighted inputs, compares the result to a threshold, and probabilistically turns on if the threshold is exceeded. The output form is the word most compatible with the set of activated output nodes. During a learning phase, the network compares its own version of the past tense form with the correct version provided by a "teacher," and adjusts the strengths of the connections and the thresholds so as to reduce the difference between the actual state of each output node and the correct state. By this process of recording and superimposing contingencies between bits of sounds of stems (e.g., the distinctive features of endings such as *-op* or *-ing*) and bits of sounds of past tense forms (e.g., *-opped* and *-ang*), the model improves its performance over time, and can generalize to new forms on the basis of their featural overlap with old ones. This model, paradigmatic of the "Parallel Distributed Processing (PDP)," "Connectionist," or "Neural Networks" approach to cognition, is frequently seen as an important piece of support for the PDP approach, and as posing a severe challenge to rule-based approaches to language and cognition (see, e.g., Rumelhart & McClelland, 1986; Sampson, 1987a; Smolensky, 1988).

Pinker and Prince (1988) challenged the psychological reality of the model on several grounds. The one relevant here is Rumelhart & McClelland's explanation of the sequence of overregularization. In building a PDP model, there are numerous ways to bias it toward conservative recording of individual input items, toward liberal overgeneralization according to frequent patterns, or some combination. The challenge was to duplicate the child's transition from conservatism to overgeneralization in a single model. They proposed a simple and ingenious hypothesis. Not only are irregular verbs high in frequency, but the opposite is true as well: the verbs highest in frequency are irregular. For example, the top ten verbs in Kucera & Francis's (1967) frequency list are all irregular. If children acquire verbs in order of decreasing frequency, they will develop a vocabulary with an increasing proportion of regular verbs as they begin to run out of the high-frequency irregulars and encounter more and more regular verbs. In particular, Rumelhart and McClelland assumed that at some point in development the child shows "explosive" vocabulary growth, which would result in a sudden influx of a large number of regular verbs. Because the regular pattern will be exemplified by many different verbs, there will be many strong links between stem features and the features defining the *-ed* ending. The effects of these newly-modified links could overwhelm the existing links between idiosyncratic features of irregular stems and the idiosyncratic features of their pasts, resulting in overregularization. As the irregulars continue to be processed, the discrepancies between the overregularized and teacher-supplied correct forms will be noted, and the crucial idiosyncratic links will be strengthened over time, eventually allowing the irregular forms to reappear.

Given these assumptions, Rumelhart and McClelland were able to model the developmental sequence with one additional assumption: The vocabulary explosion occurs after the child has just acquired his tenth verb. Their ten-verb decision results in two training phases. First, the model is presented with the ten highest-frequency verbs (excluding *do* and *be*, which can also be auxiliaries), of which only 2 (20%) happen to be regular, 10 times apiece. Then the model is presented with that list plus the 410 next-most-frequent verbs, constituting a set in which 80% of the verbs are now regular, 190 times apiece. In Phase 1 the model learned the 10 verbs successfully; when Phase 2 begins on the eleventh cycle, and it is suddenly swamped with regulars, the model overregularizes the irregulars. The recovery process begins immediately, and reaches asymptote shortly before the 200th epoch (see Figure 1, taken

from Rumelhart & McClelland, 1986).

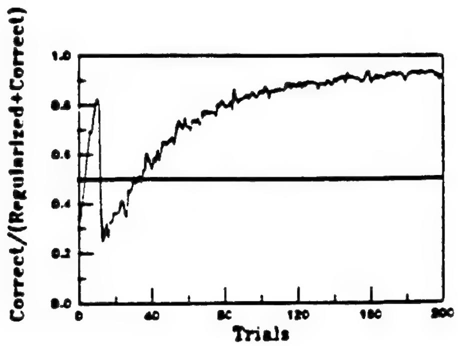


Figure 1. The performance of the Rumelhart-McClelland model on irregular past tense forms. Tendency to overregularize is estimated as the ratio of the strength of the correct irregular response to the sum of the strengths of the correct and the overregularized responses. Points below the line correspond to a tendency to overregularize. From Rumelhart & McClelland, 1986).

Pinker and Prince (1988) examined Rumelhart and McClelland's assumptions about development. Rumelhart and McClelland cited Brown (1973) in support of their assumption of a vocabulary spurt near the onset of overregularization, but Brown did not discuss vocabulary acquisition at all, and according to standard sources (see, e.g., Ingram, 1989) children's "word spurt" usually occurs at 1;6, more than a year too early to account for the onset of overregularization. Pinker and Prince examined Brown's (n.d.) vocabulary lists for Adam, Eve, and Sarah, from 5 evenly spaced samples spanning the overregularization sequence, plus a fourth child in the one-word stage. They found no explosive growth in vocabulary near the onset of overregularization, nor, more significantly, an increase in the percentage of the child's vocabulary samples that was regular: The proportion regular stayed around 50% before, during, and after the onset. Pinker and Prince also cited data suggesting that the proportion of verb tokens that are regular in parental speech to children is about 20 to 30% during overregularization, nowhere near the 80% proportion that Rumelhart and McClelland used to override the irregular patterns. They argued that a an endogenous transition from rote to rule is still required to account for the data, as in the traditional account.

Rumelhart and McClelland's model was unprecedented in the precision of its quantitative predictions about child language data. Indeed, Rumelhart and McClelland were far ahead of the field of developmental psycholinguistics in the quality and quantity of data they required for proper tests of their hypotheses. As such they have underscored the fact that there are no systematic quantitative reports of the developmental course of overregularization, its distribution across children, verbs, and time, its relation to the child's vocabulary size, and the lexical factors that cause some verbs to be overregularized more than others. This data gap has left many fundamental questions unanswered. For example, Stemberger (1989) and Marchman (1988) have questioned whether there has been adequate documentation that a U-shaped developmental sequence even exists. Maratsos (1987) questions whether there is a Blocking or Uniqueness mechanism in children, given that overregularizations and their correct counterparts coexist for years in a given child. Sampson (1987b) argued that Pinker & Prince's analysis is

misleading for sampling reasons: In small samples, verbs with low token frequencies will be underrepresented. Since regular verbs are lower in token frequency than irregulars, and newly-acquired regulars might be lowest of all, the number of regular types could be underestimated, as would their rate of growth.

2.4. Open Questions

This paper is an attempt to fill these gaps. Using the large set of transcripts of children's spontaneous speech recently made available by the ChilDES project (MacWhinney and Snow, 1985, 1990), together with the previously published tallies of children's vocabulary, we document the process of overregularization in quantitative detail, focusing on the theoretical questions described above.

The paper consists of three parts. First, we describe the overall rate and course of overregularization. We measure the overall rate of overregularization, a figure of obvious importance: one's explanation for children's behavior would be very different depending on whether they make overregularization errors 1%, 50%, or 99% of the time. We then examine whether this rate shows great variation across time, children, verbs, and combinations of these sampling units (e.g., whether the overall overregularization stage for a child is actually a composite of individual verbs each being overregularized intensively for a brief interval). In the final part of that section, we test for the existence of a U-shaped sequence.

In the second section we examine Rumelhart & McClelland's hypothesis that overregularization is triggered by an increase in the proportion of regular forms that the child processes during development. We correlate children's overregularization rate with changes in the number and proportion of verbs that are regular in their speech, the speech they hear from their parents, and their vocabularies.

In the third section we test hypotheses about the causes of overregularization by focusing on properties of different verbs that might cause them to be overregularized more or less often. We test factors related to memory strength, salience of the relatedness of stem and past forms, and complexity of the mapping from stem to irregular form.

Finally, we discuss a simple theory that accounts for most of the data. The theory maintains the traditional distinction between rule and rote, while incorporating an interesting novel feature of the Rumelhart-McClelland model.

3. Method

3.1. Subjects

The ideal data set for this study would come from a diary study recording frequently sampled (e.g., daily) conversations from several children over a span of years. Unfortunately such data are unavailable and we must use different subsamples of the ChilDES database of English-speaking children in order to address different questions.

To assess overall overregularization rates we wanted to examine as many children as possible: 69

children and 11,514 utterances containing past tense forms of irregular verbs. Table 1 shows the children, their ages, and the frequency of the transcripts. We focus on 10 children with longitudinal samples in the English ChilDES database, and 15 children with single samples from Hall and Tirre (1979). This database was augmented by small data sets from 44 more children.

Table 1
Children Studied

Child	Age	Source	Total Samples	Sampling Frequency
Abe	2;6-5;0	Kuczaj (1977)	210	weekly
Adam	2;3-5;2	Brown (1973)	55	2-3/month
Allison	1;5-2;10	Bloom (1973)	6	occasional
April	1;10-2;11	Higginson (1985)	6	occasional
Eve	1;6-2;3	Brown (1973)	20	2-3/month
Naomi	1;3-4;9	Sachs (1983)	93	weekly to monthly
Nat	2;8	Bohannon (1977)	21	within 1 month
Nathaniel	2;3-3;9	Snow	30	weekly
Peter	1;3-3;1	Bloom (1973)	20	monthly
Sarah	2;3-5;1	Brown (1973)	139	weekly
15 children	4;6-5;0	Hall & Tirre (1979)	30	2 days/child
20 children	1;6-6;2	Warren-Leubecker (1982)	20	1 sample/child
24 children	2;1-5;2	Gleason (1980)	72	3 samples/child

To answer questions about longitudinal development (including any U-shaped sequence) and vocabulary size, we need samples that begin before the onset of overregularization and continue until performance is close to adult levels. Brown's (1973) Adam, Eve, and Sarah meet this criterion; overregularizations are absent from their early transcripts and their later transcripts extend to Brown's "Stage V" in which most inflections are supplied in their correct forms more than 90% of the time.

Finally, when examining the effects of lexical factors, we analyzed overregularization rates for individual verbs from 19 children and correlated them with different properties of the verbs. Such analyses involve a tradeoff: individual children often supply too few errors to provide the wide of range of overregularization rates and the wide range of predictor variable values needed for correlational analysis, but aggregate data are in danger of displaying averaging artifacts. Therefore in the lexical analyses we seek converging results from three sources: Kuczaj's son Abe (see Kuczaj, 1976, 1977, 1978), whose numerous overregularizations and correct irregulars make him the only individual child for which across-item correlations can be interpreted with confidence; the remaining 18 individual children; and an aggregate measure that combines the overregularization rates of the 18 children.

3.2. Procedure

We tabulated all past tense uses, correct and overregularized, for all of the children, although we used slightly different methods for different children.

For Abe, the data were gathered and tabulated by Kuczaj (1976), the boy's father. Kuczaj recorded, for each month from age 2;6 - 5;0, the number of times Abe used each of sixty-six irregular verbs, the

number of times he produced the present stem of the verb with *-ed* appended (e.g. *goed* or *breaked*) and the number of times he produced doubly marked pasts in which *-ed* was added to the irregular past form, for example *wented* or *broked*. The list includes 66 irregular verbs that Abe used in the samples, excluding *had, was, were,* and *did* which are possible auxiliaries, and verbs like *hit* and *put* that do not differ in their stem and past forms (see the Appendix of Pinker and Prince, 1988, for an exhaustive list of irregular verbs in present-day American English, sorted into subclasses.)

For Adam, Eve, and Sarah, verb usages were tallied on a DEC Microvax II running UNIX. Individual transcript files were combined into single master file for each child. The "freq" program in the CLAN software package (MacWhinney and Snow, 1990) counts the number of times every word is used in a particular transcript session, for a particular speaker, and it was run on the individual transcript files and the combined files. The combined frequency list for each child was then edited to include only words that the child may have used as a verb, including words that occur only infrequently as a verb (e.g., *fish, color, ground, milk*). For all such items (i.e., all words that are not obviously exclusively verbs in the child's vocabulary), the Unix utility 'fgrep', which finds matches of regular expressions, extracted all the transcript lines they occurred in. Each of the resulting lines was checked by hand, and excluded if the matched word turned out not to be used as a verb. If a word appeared in a single-word utterance, it was excluded; thus, "sleep" or "put" appearing alone were not counted as verbs, but "Adam sleep" or "put Mommy" were included. Verbs repeated in successive sentences such as "I falled down. I falled down, Momma" were counted separately, since children are capable of saying both a correct and incorrect form of a single verb in successive utterances, as in Abe's "Daddy comed and said 'hey, what are you doing laying down?' And then a doctor came...".

As mentioned, no-change verbs present a problem because correct past tense usages are identical to stem forms and hence cannot be extracted by machine. However it would be desirable to include these forms for the longitudinal analyses and so for Adam, Eve, and Sarah, we searched for all instances of the verb and checked the original transcripts to tally correct past usages. Where the transcript did not provide information regarding verb tense, the context was used. Contractions such as *gimme, gonna, I'm, it's,* or *doesn't,* were excluded, as were participles such as *broken* or *gone,* and the quasi-auxiliary *used to.* All of the usages of *have, be,* and *do* as main verbs were included, as were some of their usages as auxiliaries; most tokens included were main verbs (see Stromswold, 1990, for a study of the overregularizations of main verb versus auxiliary usages of these verbs). A very small number of mimicked utterances at early ages, regular participles, and irregular participles that are identical to past tense forms may have been included. Intentionally included were verbs that were not very clearly uttered, but were clear enough for the transcriber to have made a reasonable guess, and some slight phonetic variations such as *-in* for *-ing,* particularly for Sarah whose samples were transcribed more narrowly than the others. (However Brown, 1973, notes that all the speech was carefully transcribed with regard to presence or absence of phonetic material corresponding to inflections.)

For the other children, and the Gleason and Warren collections, we used a Sun Microsystems Sparcstation 4 running under UNIX to tabulate all irregular past tense utterances. Using the 'freq' program we extracted the number of occurrences for each irregular verb listed in the Pinker & Prince Appendix, together with all forms that ending in *-ed*. We then collected all the overregularization errors from this list, by removing regular verbs and other part of speech categories. Like the data for Abe, but

unlike those for the Brown children, no-change verbs and the auxiliaries (*did, was, were, had*) were excluded. Due to the lack of hand-checking for this large collection, we were unable to exclude overregularized participles such as in *the window was broked*. Since we did exclude correct participle forms if they were distinct from the past tense forms (about 60 irregular verbs have this property; see Pinker & Prince, 1988), this can result in an overestimate of overregularization rates, though it would be quite small. The word *seed* presented particular problems since most of its uses are as a noun rather than as an overregularization of *see;* these were eliminated by hand. Though *read* is not phonetically a no-change verb, orthographically it is, and it was excluded from the analyses. Repetitions were counted separately, as with the Brown children.

The verb *get* is problematic. When adults speaking the standard dialect use *got* in a stative possessional sense, as in *I've got an ice cream cone*), it is the perfect participle of *get* (meaning "obtain"), accompanied by some form of the auxiliary *have*. The meaning is possessional because of the semantics of perfect aspect in English: if the state resulting from obtaining something in the past currently holds, you possess it now (Bybee, 1985). However, if children do not attend to the auxiliary, it would be natural for them to reconstrue *got* as a present tense form meaning "possess," and there are numerous forms like *Look, I got an ice cream cone* that suggest they do often use *got* as a present tense verb. For these usages we would erroneously credit the child with the correct past form of *get*. Kuczaj (1976) noted this problem and used the context to distinguish present from past usages of *got* in Abe's speech. For the other children, unfortunately, our observed overregularization rates for *get* are probably underestimates, and since *get* is a frequent verb, the overall overregularization rates across verb tokens will be, too. The degree of underestimation is small, however: if all forms of *get* are omitted from the tallies, the estimated overregularization rates do not systematically change by more than a percentage point or two.

To verify that these machine-generated tabulations were likely to be accurate, we compared an exhaustive hand tabulation with machine-generated totals for a single set of transcripts, Abe's. The hand-generated totals were calculated by (1) extracting all utterances containing irregulars and all utterances with forms containing *-ed* and (2) checking these utterances to remove all participles, nouns, and all other nonpast forms. We compared the overregularization rates derived by the two methods across the 66 verbs Abe used. The two sets of estimates correlated strongly (r = .81). Most of the discrepancy can be attributed to small samples for some verbs, where disagreement over a single sentence can greatly affect the overregularization rate for that verb (e.g., if a verb was used twice, once correctly and once incorrectly, its estimated overregularization rate can vary from 0% to 100% if one of the sentences is omitted or misclassified). When only those verbs that were used a minimum of 4 times were included, the machine-generated and hand-tallied rates correlate .90, and for verbs used a minimum of 10 times, the rates correlate .98. Thus in several analyses in this paper we will exclude all verbs used less than a certain minimum number of times, usually 10. Secondly, we calculated the overall average of overregularization rates for 66 verbs; the difference between the averages for the two methods was less than half a percentage point.

In addition, we tabulated the irregular past tense utterances of the four sets of parents who used the most past tense forms, those of of Abe, Adam, Sarah, and Peter.

We defined "overregularization rate" as the proportion of tokens of irregular past tense forms that

are overregularizations; that is,

$$\frac{\text{\#overregularization tokens}}{\text{\#overregularization tokens + \#correct irregular past tokens}}$$

Overregularizations included stem+*ed* forms like *eated* and past+*ed* forms like *ated*. Virtually all of children's past tense forms are in past tense contexts (Brown, 1973; Kuczaj, 1976), so there is no need to take into account the semantic correctness of the tense marking. Overregularization rates were calculated over tokens for a given verb for a given child, over tokens of all verbs for a given child, and over tokens of all verbs for all children.

An alternative measure, used by Stemberger (1989), is the proportion of past tense tokens of irregular verbs that are incorrect, or

$$\frac{\text{\#overregularizations + \#stems in past contexts}}{\text{\#overregularization tokens + \#stems in past contexts + \#correct past tense-form-tokens}}$$

For calculations across verbs, these two measures are virtually identical because once children begin to overregularize they rarely use the base form in past tense contexts (Kuczaj, 1977), an effect we will review in more detail. But for testing developmental trends the two measures are very different for early stages in which child usually use the stem form in past tense contexts. Overregularization rate, rather than error rate, is the appropriate measure for the hypotheses under investigation here, because error rate confounds two separate issues. The marking of past tense is not innate, and children take some time to begin doing it reliably (Brown, 1973). Overall error rate confounds the child's mastery of *whether* to mark a verb in the past tense with *how* the child marks a verb for past tense given that he or she has decided to mark it at all. While it is possible that children who use a base form in a past context are mentally representing it as a no-change irregular past form akin to *hit* and *put*, there is no good reason to credit the child with that hypothesis; it is more parsimonious to assume that the child is not yet using inflection at all and is just uttering the stem form, as he or she does for progressive contexts, third-person-singular contexts, and so on (Brown, 1973). Since the issues discussed in the Introduction concern *how* the child chooses one or another past tense form (for example, Rumelhart and McClelland assume that the past/nonpast tense distinction has independently been mastered and simply feed correct stem-past pairs to their model), the proportion of past tense irregular forms that are overregularizations is the suitable measure.

Note finally that our study excludes two kinds of errors that are sometimes lumped with overregularization but that involve different phenomena. In languages with richer inflectional systems than English, children often inflect a stem with an affix that is incorrect for some feature of the stem, such as gender or animacy, or the context, such as case or definiteness (Slobin, 1973; Pinker, 1984). Such errors are best characterized as *underdifferentiation* -- an insensitivity to some systematic distinction relevant to inflection -- rather than overregularization, which involves idiosyncratic lexical exceptions to a systematic inflectional process. Second, we are excluding overapplications of *irregular* patterns in errors like *tooken* or *brang*. According to the hypothesis we are considering in this paper, they would represent competition between two irregular forms for one memory slot (see Ullman & Pinker, 1990, for discussion), not a competition between a stored irregular and the regular rule. The distinction is not important for present

purposes; Bybee and Slobin (1982a) point out that irregular overirregularizations are extremely rare in preschool children's language, and we found very few in our samples.

4. The Rate of Overregularization

The first question to address is how frequent overregularization is. As mentioned, the most frequent characterization – despite the fact that it was proven false more than 25 years ago – is that once children begin overregularizing they do so all the time, replacing correct irregulars altogether. More informed characterizations have the child varying freely between overregularized and correct irregular forms, suggesting that the child fails to discriminate between the two, with a preference for the irregular emerging slowly and culminating in adult performance. Such a stage of free variation calls any kind of Blocking or Uniqueness principle into question (Maratsos, 1987), and leaves unsolved the learnability problem of eliminating the incorrect forms. But a third empirical pattern is possible – that overregularization errors are rare relative to correct irregulars. If so, there would be no qualitative difference between children and adults. Both would discriminate between regularized and irregular forms, presumably because application of their regularization mechanism to listed irregulars is blocked. Children's occasional overregularizations, like adult's speech errors and their uncertainty about low-frequency irregulars like *hove*, could be attributed to probabilistic imperfect retrieval from rote memory. Strictly speaking, it would suffice to show that the overregularization rate is reliably less than 50% to establish that children possess the knowledge that the irregular past tense form of a verb has priority over the regularized form, but of course such a claim would be more convincing the lower the overregularization rate turns out to be.

4.1. Overall Overregularization Rate and Distribution Over Children

We first calculated overall overregularization rates. The median overregularization rate across the 25 children with individual transcripts was 2.5%. Table 2 shows the relevant data for the different children, whose distribution of overregularization rates are plotted in the histogram in Figure 2. The distribution is roughly exponential, with most children at the extreme low end; only three children, Naomi, April, and Abe, overregularized more than 8% of the time (9%, 13%, and 23%, respectively.) If we exclude Abe, who is both the extreme outlier in overregularization rate and the contributor of the largest number of tokens, then 3.1% of all irregular past tokens from the 68 children were overregularized (7.0% including Abe), and the average overregularization rate across the 24 children with their own transcripts children was 3.4% (4.2% including Abe). Thus, the global data suggest that overregularization is a relatively rare phenomenon; if they legitimately reflect children's tendencies, it would suggest that children's language systems, like adults', are strongly biased to suppress overregularization, contrary to common belief.

Table 2
Overregularization Rates for Individual Children

Child	Correct	Stem+*ed*	Past+*ed*	Total	Overreg Rate
Abe	1676	416	95	2187	0.23
Adam	2774	44	5	2823	0.02
Allison	31	2	0	33	0.06
April	47	6	1	54	0.13
Eve	302	23	1	326	0.07
Naomi	378	34	2	414	0.09
Nat	52	0	0	52	0.00
Nathanie	1243	11	3	257	0.05
Peter	853	17	4	874	0.02
Sarah	1993	60	4	2057	0.03
Hall children:					
ANC	79	2	0	81	0.02
BOM	112	1	0	113	0.01
BRD	128	2	0	130	0.02
CHJ	151	4	0	155	0.03
DED	106	5	0	111	0.05
KIF	100	0	0	100	0.00
MIM	77	0	0	77	0.00
TRH	47	3	0	50	0.06
VOH	64	1	0	65	0.02
GAT	159	10	0	169	0.06
JOB	130	0	0	130	0.00
JUB	132	8	0	140	0.06
MAA	105	2	0	107	0.02
TOS	84	0	0	84	0.00
ZOR	98	0	0	98	0.00
Aggregate Databases:					
Gleason	472	32	1	505	0.07
Warren	317	4	1	322	0.03
Total:	10710	687	117	11514	0.07

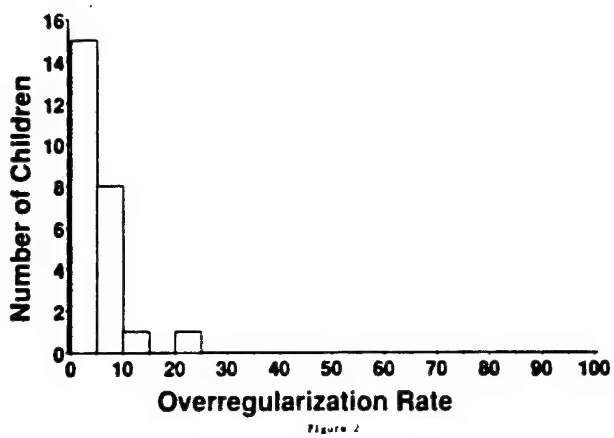


Figure 2. Histogram of overregularization rates across 25 children.

4.2. Distribution of Overregularization Rates Over Time

Of course, the surprisingly low rates obtained may be an averaging artifact: each child might go through a circumscribed U-shaped period of indiscriminate overregularization, preceded and followed by many more months of near-perfect performance. For example, if there were a span in which overregularization did not occur for 28 months, but went up to 96% for one month, the average overregularization rate for the span would only be 3.3%. This possibility is shown in Figure 3. (In all our developmental graphs, we plot percentage correct (1 - overregularization rate) rather than percent overregularized, so that regressions in development appear as U's, not inverted U's.) Figures 4-6 plots the overregularization rates over time (pooled across one-month collections of 2-4 samples) for the children studied longitudinally (Adam, Eve, and Sarah). They demonstrate that low rates characterize the entire period of overregularization. Adam's worst month for overregularization had a rate of only 6%; Eve's was 21%; Sarah's was 12%; and because these figures are maxima selected post hoc they may overestimate the worst true overregularization rate because of sampling error. Even Abe, an extreme outlier among 25 children, overregularized only 50% of the time in the month of samples selected as the worst (see Figure 7).

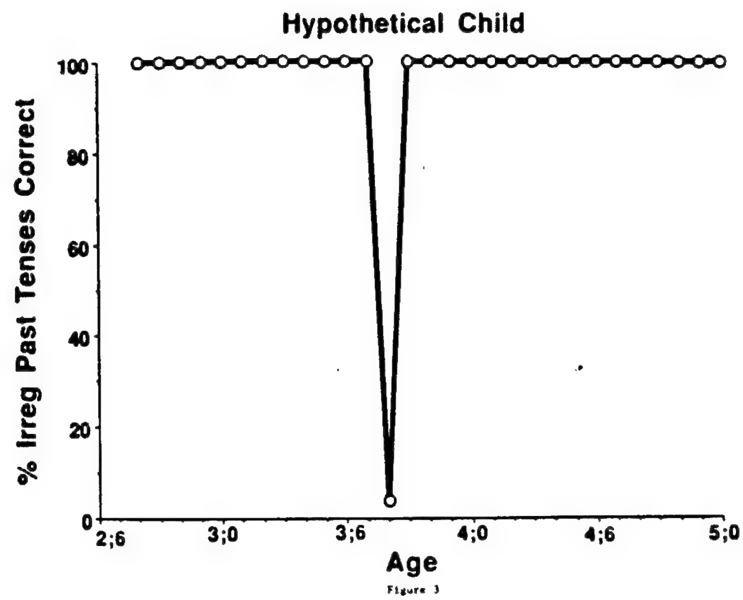
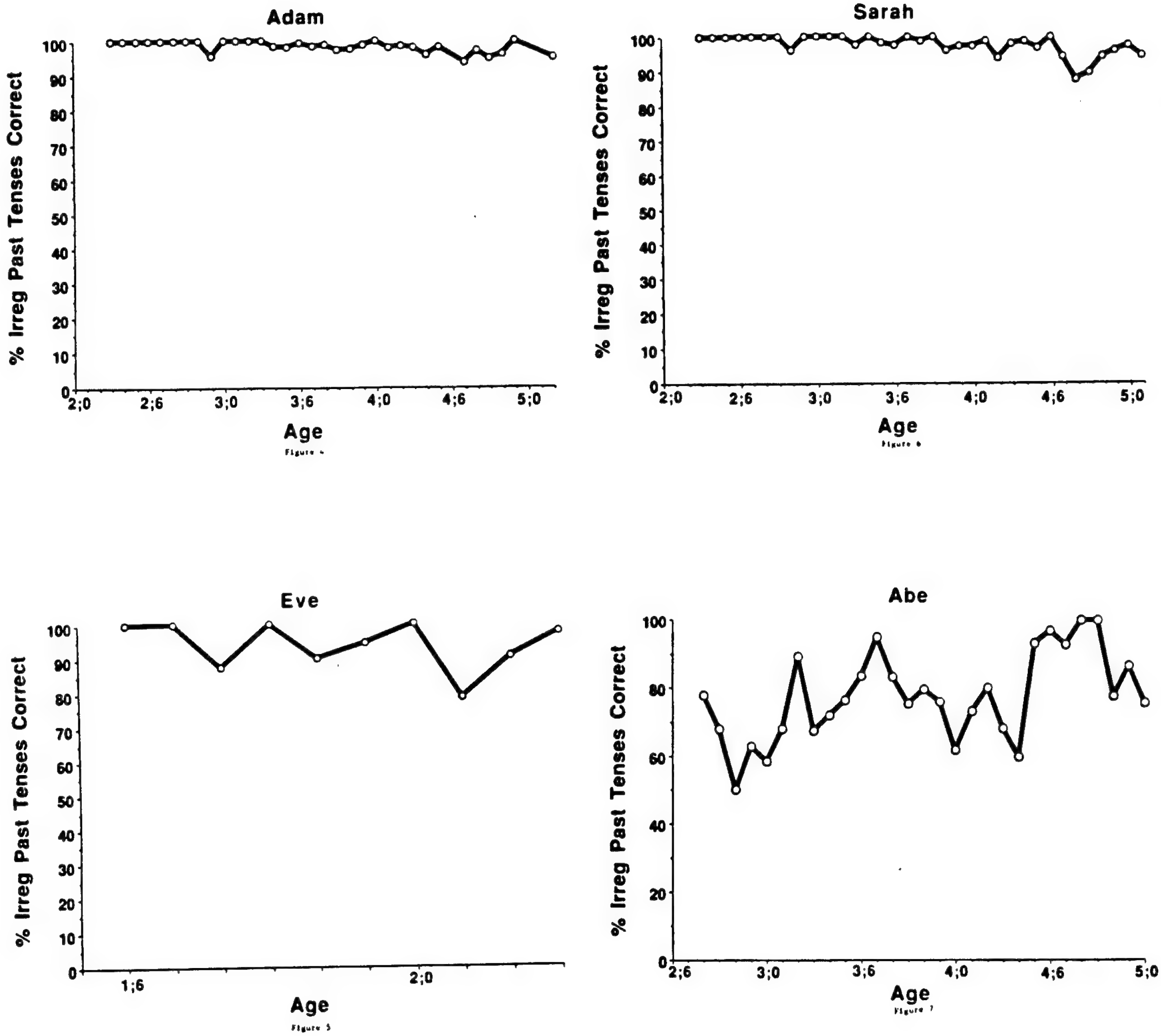


Figure 3. Hypothetical developmental sequence that would yield low overregularization rates as an averaging artifact.



Figures 4-7. Overregularization rates of Adam, Eve, Sarah, and Abe at different ages.

Furthermore for all four children overregularization begins early in the span sampled and lasts for the rest of the period: ignoring Eve (whose samples end at 2;3) we see fairly steady overregularization from some time in the third year at least into the sixth year. Only for Abe is there any evidence for a reduction in overregularization rate late in development.

4.3. Distribution of Overregularization Rates Over Verbs and Children

Another possibly misleading effect of averaging involves individual verbs: Perhaps many verbs are never or rarely overregularized, but some are overregularized most of the time or at least indiscriminately. Because these verbs may differ from one child to another, the question can only be addressed by examining overregularization rates for different verbs in individual children. However, sampling error can make such counts extremely misleading. In the extreme case, if a verb is only used once, its observed overregularization rate can only be either 0% or 100%, regardless of the child's actual overregularization tendency; if it is used twice, the observed rate can be 0%, 50%, or 100%, and so on. Even with a low overall regularization rate and slightly larger samples, high estimates of overregularization rates for a given verb frequently will arise by chance. If we restrict attention to only those irregular verbs that were used 10 times or more in the past tense, we find that Adam did not overregularize any of his 32 verbs at a rate higher than 10%, Eve overregularized *fall* 8 out of 10 times but did not overregularize any of the other 9 more than 20%, and Sarah overregularized *throw* 7 out of 10 times but did not overregularize any of the other 25 more than 30%.

Abe is the only child who overregularized often enough for us to examine a large number of irregular verbs statistically. The overregularization rates for the 66 irregular verbs he used are listed in Table 3. Though the overregularization rates range from 0 to 100, this by itself this does not show that verbs are overregularized at different rates. Even if Abe's overall overregularization rate of 23% applied uniformly to every verb, high or low observed values for particular verbs can occur with reasonably high probability, especially with small samples sizes. Thus a somewhat less misleading picture of the distribution of Abe's verbs can be seen in Figure 8, a histogram of overregularization rates for different verbs, showing only those verbs that were used 10 times or more.

Table 3
Overregularization Rates of Individual Verbs

	Abe					Aggregate	
	Correct	Stem+*ed*	Past+*ed*	Overreg Rate	Log Adult Frequency	Standardized Overreg Rate	Log Adult Frequency
bend	1	1	0	0.50	1.10	0.53	0.51
bite	1	4	0	0.80	3.33	-0.43	2.97
blow	4	7	0	0.64	0.00	–	0.51
break	32	4	12	0.33	2.77	-0.30	3.38
bring	2	5	1	0.75	2.48	-0.57	2.71
build	1	8	0	0.89	1.10	–	0.85
buy	7	5	0	0.42	1.61	-0.38	2.96
catch	21	5	1	0.22	2.64	-0.47	2.35
choose	1	1	0	0.50	0.00	–	–
come	19	50	4	0.74	3.85	-0.30	3.41
dig	1	3	0	0.75	0.00	–	–
draw	1	6	0	0.86	1.95	2.11	1.25
drink	5	5	0	0.50	2.40	–	1.45
drive	0	2	0	1.00	1.39	–	0.69
eat	82	18	2	0.20	4.03	-0.46	3.09
fall	72	49	2	0.41	3.71	0.10	3.36
feed	0	1	0	1.00	0.00	–	0.41
feel	5	10	0	0.67	1.79	–	1.79
fight	1	2	0	0.67	1.79	–	1.25
find	142	1	2	0.02	3.99	-0.36	3.14
fly	4	4	0	0.50	1.61	1.30	1.39
forget	67	0	0	0.00	3.22	-0.41	2.94
freeze	1	0	0	0.00	0.00	–	0.00
get	244	4	48	0.18	5.51	-0.40	5.75
give	5	2	0	0.29	2.71	-0.33	3.43
go	113	53	4	0.34	4.79	-0.25	4.38
grow	2	6	0	0.75	1.10	–	1.10
hang	0	4	0	1.00	2.08	–	1.50
hear	11	27	0	0.71	3.09	-0.21	2.67
hide	2	0	0	0.00	0.69	–	0.69
hold	0	4	0	1.00	1.10	–	0.69
keep	1	0	0	0.00	1.10	–	1.39
know	11	6	0	0.35	2.83	-0.55	2.48
leave	13	2	0	0.13	3.64	-0.50	3.37
lose	8	1	0	0.11	2.89	-0.45	3.20
make	147	23	3	0.15	4.62	-0.32	4.20
mean	3	1	0	0.25	1.79	–	1.56
read	1	1	0	0.50	4.06	–	4.06
ride	1	1	0	0.50	0.00	-0.60	1.18
run	4	6	0	0.60	2.08	0.47	1.87
say	277	3	0	0.01	4.76	-0.40	4.63
see	158	3	2	0.03	4.66	-0.31	4.01
send	8	1	0	0.11	1.95	–	1.50
shake	2	0	0	0.00	0.00	–	0.00
shoot	10	2	1	0.23	2.08	-0.60	1.79
sing	2	1	0	0.33	0.69	–	0.69

sit	1	1	0	0.50	0.69	-0.48	1.79
sleep	1	2	0	0.67	1.10	–	1.39
spend	0	1	0	1.00	1.10	–	0.69
spit	0	2	1	1.00	0.00	–	–
stand	1	2	0	0.67	0.00	–	0.29
steal	3	0	0	0.00	1.79	0.32	1.25
stick	1	1	0	0.50	2.40	0.05	2.51
string	0	1	0	1.00	0.00	–	0.00
sweep	0	2	0	1.00	0.00	–	0.00
swim	0	2	0	1.00	0.00	–	0.00
swing	0	3	0	1.00	0.00	–	–
take	21	6	3	0.30	3.64	-0.33	3.64
teach	3	0	0	0.00	1.39	-0.55	2.01
tell	52	7	0	0.12	3.93	-0.42	3.70
think	75	9	9	0.19	4.91	-0.57	4.36
throw	5	12	0	0.71	1.95	1.09	1.95
wake	3	1	0	0.25	1.10	1.67	0.92
wear	0	3	0	1.00	1.10	–	0.98
win	16	12	0	0.43	2.83	0.77	2.34
write	1	7	0	0.88	1.39	1.01	1.85

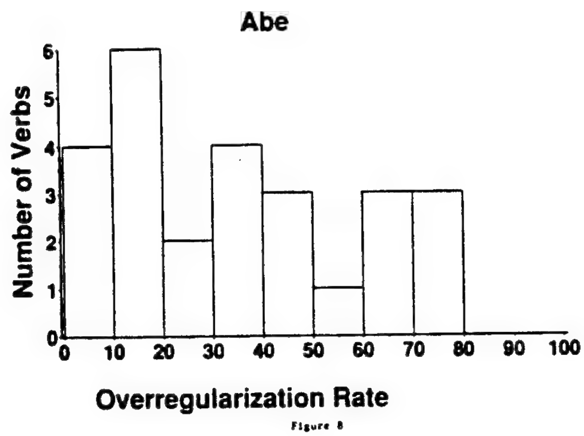


Figure 8. Histogram of overregularization rates of Abe's verbs (10 or more tokens per verb).

We performed the following test of the null hypothesis that all of Abe's verbs had a single probability of overregularization of .23; that is, whether there are more verbs with higher rates than .23 than would be expected by chance, given their sample sizes. Using the normal approximation to the binomial distribution, we calculated a *z*-score for each verb corresponding to the probability that one would obtain the observed proportion of the tokens overregularized or greater if the underlying probability was .23. We then partitioned the range of probabilities into 5 bins, corresponding to equal expected numbers of verbs per bin under the null hypothesis of a constant overregularization rate, and compared the observed numbers of verbs in each bin. The observed and expected frequencies differed significantly, $X^2(4) = 82.79$, $p < .0001$. Since the normal approximation to the binomial may be inaccurate for small samples, we replicated the analysis excluding all verbs that Abe used fewer than 10 times; the 26 verbs, sorted into five bins of probabilities of overregularization rate, differed significantly from what would be expected under the null hypothesis of a constant overregularization rate, $X^2(4) = 20.92$, $p < .0005$. In sum, Abe did overregularize different verbs at different rates and estimates of overall overregularization rates must be qualified by examining different verbs; across-verb correlational analyses are reported in Section 7. Nonetheless most of Abe's verbs, like those of Adam, Eve, and Sarah, fall at the low end of the distribution of overregularization rates when misleadingly small samples are excluded; the overall low overregularization rate is not an artifact of averaging, say, a few verbs that are always overregularized with many verbs that are never overregularized.

4.4. Distribution of Overregularization Rates Over Words, Time, and Children

The most stringent test of the hypothesis that overregularization is a probabilistic and relatively rare event would look at at the fate of individual irregular verbs for individual children as they grow older. This is independent of the overall level of overregularization for different verbs that we have just examined, just as the waveform of a sound wave is is independent of its amplitude and of its DC component. For example, it is possible that each child goes through a stage for each verb when the verb is overregularized exclusively (see Figure 9), or as often as it is produced correctly. If these stages are fairly brief and circumscribed, the steady low rate of regularization could be an averaging artifact of a sequence of deep narrow U's, one for each verb. If so, or if the verbs all follow some other identical developmental curve, but not in phase, the protracted period of overregularization would reflect the application of blocking to different verbs at different times.

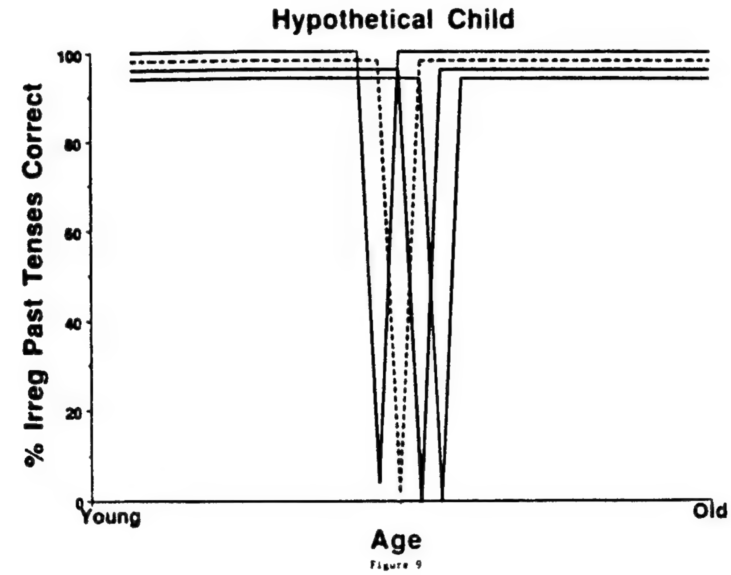


Figure 9. Hypothetical developmental sequence for different verbs that would yield low overregularization rates as an averaging artifact.

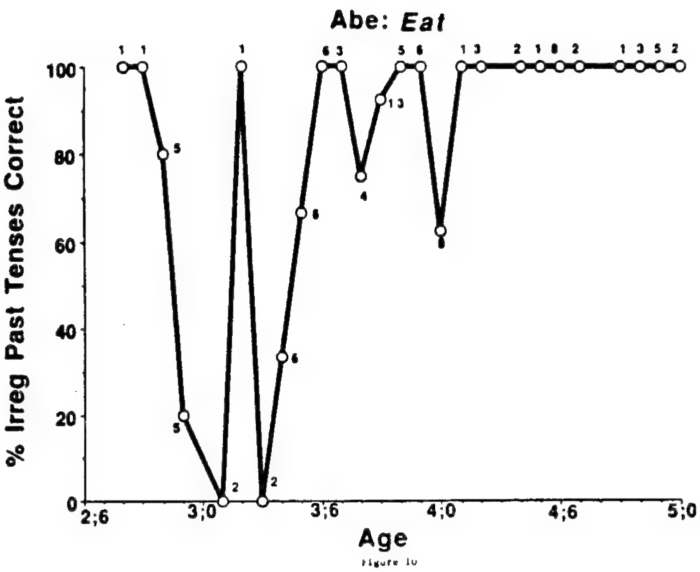
There are other possible interactions among children, verb, and age that would noteworthy. One can see whether any verbs cease to be overregularized altogether before the end of the period, or alternatively whether children begin to stop overregularizing all verbs at the same time; the latter finding would suggest that the child learns or develops the blocking principle only at that point. Another possibility is that different verbs follow different and largely unsystematic patterns, perhaps because, as we have suggested, overregularization is a quasirandom performance deficit.

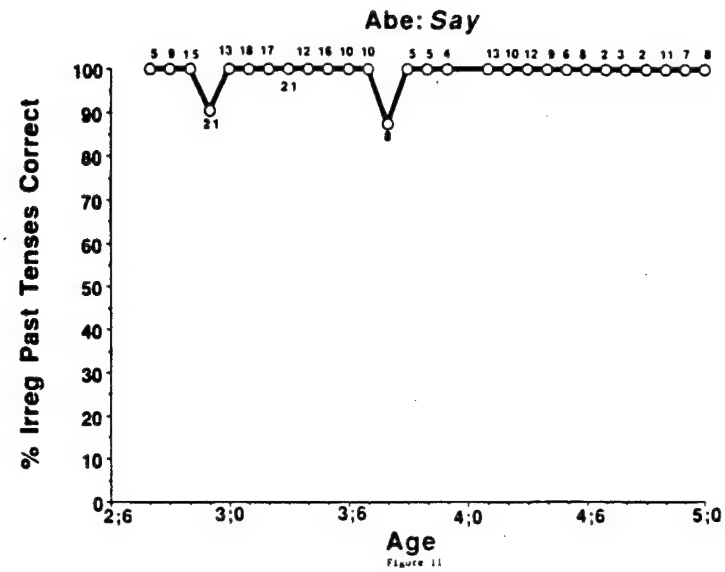
Clearly low rates of overregularization are not an artifact of a sequence of transient overregularization stages, one for each verb: individual verbs can be overregularized across large spans of time. The very first verb that Adam overregularized, *feel* at 2;11, he also overregularized in his last sample at 5;2. Similarly, *throw* was overregularized at 3;4 and at 4;4; *make* at 3;5, and 5;2; *fall* at 3;5 and 4;10. Sarah's first overregularization, *heared* at 2;10, appeared again at 4;11; *winned* and *maked* also made appearances in the samples separated by a year or more. Even the 9 months' worth of samples from Eve contain *falled* at 1;10 and again at 2;2. Likewise more than half of Abe's overregularized verbs (38) were overregularized over a span of one year or more; 24 were overregularized over a span of two years or greater.

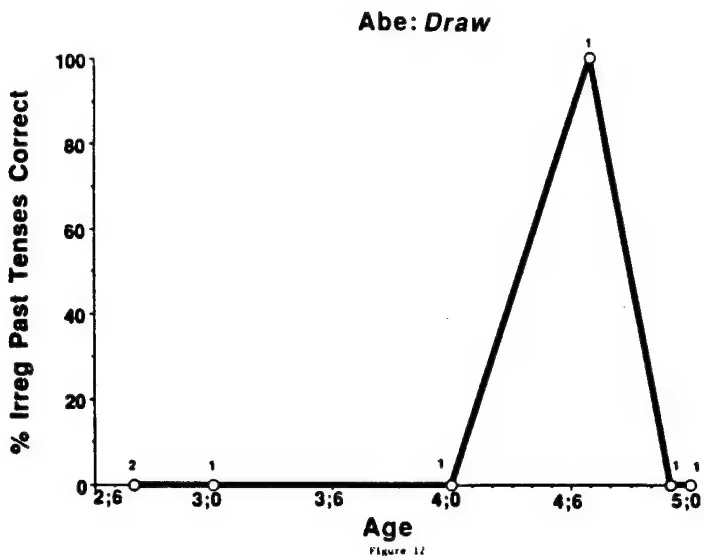
Unfortunately, when we turn away from the simple question of whether overregularizations of a given word reappear across long time spans, and try to trace each one of a child's irregular verbs over time, we run up against severe sampling limitations. As mentioned, small samples frequently yield

inaccurate estimates of overregularization rates, and many of the samples of tokens of a given verb for a given child in a given month were very small. Thus, developmental curves for individual verbs with low token frequencies for the child can oscillate wildly among a few discrete values, revealing little about changes in the underlying true rates. With these caveats in mind, we now examine curves for individual verbs for Abe, the biggest overregularizer.

For most of the 66 irregular verbs Abe used, the curves can best be described as chaotic, and highly variable from verb to verb. They are most conveniently summarized by grouping them into four kinds of patterns, shown in Figures 10 through 13. Some verbs, like *eat* (Figure 10), are overregularized in the earlier transcripts but appear to be completely mastered before the end of the sampling period. (Other verbs with this pattern were *bite, break, catch, fall, go, make, think,* and *throw.*) A second class of verbs, such as *say* (Figure 11) are rarely overregularized at any point. (*Find, forget, see,* and *tell* are the others.) A few, such as *draw* (Figure 12) are overregularized throughout the sample (*build* is similar); however, such verbs were used only rarely, and the high observed rates could in part be due to sampling error. But many verbs, such as *win* (Figure 13), *blow, buy, come, feel, get, know,* and *shoot*, show no interpretable trend, oscillating among periods of high and low overregularization.







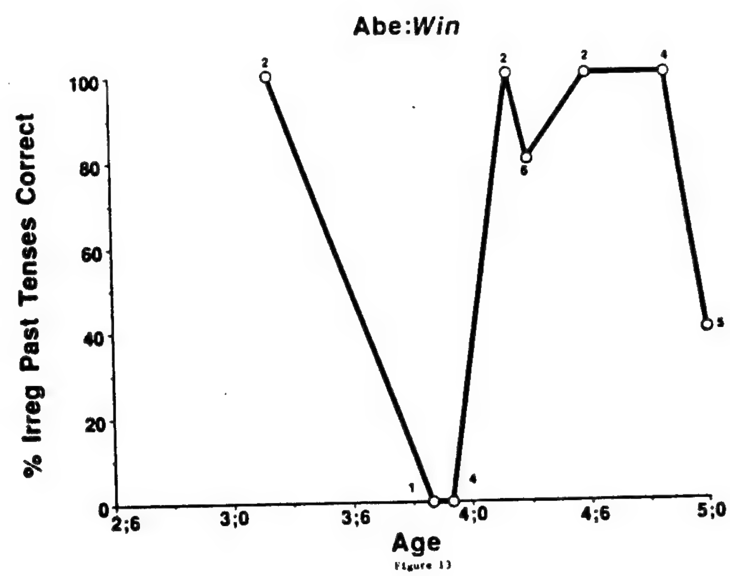


Figure 10. Example of a verb that resists overregularization as the child gets older.

Figure 11. Example of a verb that is rarely overregularized at any age.

Figure 12. Example of a verb that is overregularized throughout development.

Figure 13. Example of a verb with a chaotic developmental pattern.

In sum, apart from haphazard variation (possibly due to sampling error) and overall low or high rates (to be discussed in Section 7), the only meaningful temporal pattern for individual verbs is that some appear to be mastered in the late 4's, resisting further overregularization.

4.5. When Does Overregularization Cease?

There are no signs of overregularization going away or even decreasing in Adam, Eve, or Sarah's samples, which last through the early 5's for the former and latter. Abe improved slightly overall by 5, and ceased overregularizing a handful of verbs. Clearly, overregularization diminishes extremely gradually; Kuczaj (1977), based on a personal communication from Slobin, notes that it is still present in school-age children of 9 or 10.

Two studies provide us with estimates of the overregularization rate in older children. Moe, Hopkins, & Rush (1987) report a sample of 10,530 irregular past tense utterances from over 329 first graders. The overregularization rate in their data is 2.8%.

Carlton (1947) reports 2196 past tense tokens among the speech of 96 fourth-graders she recorded. These included thirteen overregularizations. If approximately 75% of past tense tokens in are irregular (see Sections 6.2 and 6.3), the overregularization rate for this group is 0.8%.

Do overregularization errors ever completely disappear? Joseph Stemberger has kindly provided us with the full set of past tense overregularizations in his corpus of 7,500 adult speech errors. The list includes 25 past tense overregularizations (18 of the stem, 7 of the irregular past form). In Stemberger (1989) he suggests that the rate of adult speech errors might be about one error per thousand sentences. If we assume that all sentences contain verbs, that about 10% of verbs in casual speech are in the past tense (Smith, 1935), and that 75% of adults' verb tokens are irregular (Slobin, 1971; see Sections 6.2 and 6.3) we get a very crude estimate of adults' overregularization rate of .00004 -- three orders of magnitude lower than preschoolers', and two orders lower than 4th graders.

So although in one sense both children and adults overregularize, there is also a dramatic difference in their rates of doing so. Perhaps the difference is just a consequence of hearing more tokens of each irregular verb as one lives longer, with more exposures leading to more reliably accessible memory traces. For example, a negative exponential learning curve with a time constant of one order of magnitude of improvement in retrieval probability for every five years' worth of irregular past tense tokens could handle the reported overregularization rates from the preschool years through adulthood comfortably. In the absence of more plentiful and finer-grained data, it is premature to claim that there is no qualitative difference between children and adults, but current evidence does not demand that there be a difference.

4.6. Comparison to Previous Estimates in Spontaneous Speech

Given the stereotype that children go through a stage in which they always overregularize, or even overregularize in free variation with correct irregular forms, our finding of an almost uniformly low rate of overregularization across children, verbs, and stages comes as a surprise. Why has the low rate not been noted before?

One reason is that few investigators have counted the number of correct irregular past tense uses; only the errors are reported. And when actual error rates are calculated, they have come from fairly small samples, often broken down by child, month, and verb type. In such cases rates can appear higher either because of sampling artifacts, special attention to the hardest verbs from the most error-prone children at their worst stages, or both. Kuczaj (1977) reports overregularization rates of 14 children from 2;6 to 5;6. Six of the children had rates comparable to those reported here (1.1% to 8.33%); for the other eight the rate ranged from 26.1% to 40.2%, with an average over all the children of 20.9% (though even this rate, note, is nowhere near the indifference rate of 50%.) However, their corpora consisted of only 6 to 162 irregular past tense tokens per child, with a mean of 79, far fewer than we analyze here and small enough to result in sampling overestimates. The other data his studies made famous came from his son Abe, but as we have seen, Abe happens to be the outlier from among the 25 individual English-speaking children we studied in the ChilDES database. Bybee and Slobin (1982a) report figures from a pooled sample of 31 children; they are reported separately for different subclasses of irregular verbs, which range from 10% to 80%. Unfortunately, the number of verb types and tokens in each class is not reported, and it is not clear whether the averages are calculated over verb types or verb tokens within the subclass, so overall overregularization rates cannot be computed from the data. But Slobin (1971) published actual token frequencies from 24 of these 31 children (those from the Miller and Ervin, 1964, samples. From his table one can calculate that those children overregularized at a rate of 10.2%, far closer to our estimates. Finally, it is conceivable that in some studies, children and samples with high rates of overregularization were systematically over-represented in order to provide a larger database for comparing verbs and verb types, just as we selected Abe for that reason in our investigation of lexical effects in Section 7.

4.7. Previous Estimates of Overregularization from Elicited Production Experiments

Several experimental studies have elicited past tense forms in sentence completion tasks using existing irregular English verbs. For example, in the experiment by Kuczaj (1978), the overregularization rates for the groups of 3-4 year-olds, 5-6 year-olds, 7-8 year olds were respectively 29%, 49% (42% stem overregularizations, 7% past+*ed* overregularizations), and 1%. Bybee and Slobin (1982a) found that their third-graders (8;6 - 10;1) overregularized between 2% and 55%, depending on the verb subclass. Marchman (1988) found the following overregularization rates for her different age groups: 4's, 32%; 5's, 33%; 6's, 22%; 7's, 10%; and 9's, 5% (calculated from her Table 1, based on 76% of the test items being irregular, as mentioned in her text.) Note that in all such studies the overregularization rates are generally far less than 50%; once again there is virtually no evidence that young children overregularize exclusively or in free variation with correct irregular forms.

Of course the overregularization rates in elicited production tasks are still far higher than those obtained from spontaneous speech, but the two kinds of estimates are not comparable. Bybee and Slobin (1982a), Stemberger and MacWhinney (1986), and Prasada, Pinker, and Snyder (1990) found that adults, when put under time pressure, are prone to making overregularization errors at even higher rates than children (from 6% to 31% of the time in the Bybee-Slobin study, depending on the subclass), presumably because of a greater likelihood of retrieval failure. It is plausible that many children feel that they are under some such pressure in experiments even if it is not explicitly stated. Furthermore if children ever fall into a strategy of treating each experimental item as a pure sound, rather than as a word they know, it essentially becomes a nonce form and regularization is the most accessible option.

But most important, in all such tasks children are being supplied with the stem itself seconds before they are asked to supply the past form (e.g., "This is a girl who knows how to *swing*. She did the same thing yesterday. She ____"). This contrasts with naturalistic settings in which children produce a past form for an irregular in response to a mental representation of the verb's meaning plus the feature for past tense; the phonetic form of the stem need never be activated. If the child's representation of the stem form is primed by its appearance in the elicitation task, the stem could become unusually available for the regular inflection process, and relatively less liable to being blocked by the irregular (see Stanners, Neiser, Hernon, & Hall, 1979, and Fowler, Napps, and Feldman, 1985, for evidence that stem forms and regularly inflected forms prime each other).[2] In support of this hypothesis, Kuczaj (1978) notes that past+*ed* errors, which were not uncommon in his subjects' speech and which they frequently judged as grammatical, were virtually never uttered in a sentence completion task, presumably because the stem form provided in the task intruded into the inflection process.

To test whether hearing a stem form elevates children's tendency to overregularize, we capitalized on the nicely-matched data sets made available by Abe's convenient habit of alternating between correct and overregularized forms of a given verb within a single sample. We counted the proportions of irregular and overregularized utterances within these 91 sets (each representing 1 or more overregularizations and 1 or more correct irregular versions of a given verb in a given sample) that had been preceded in the conversation by an adult utterance containing the stem form. By focusing on such sets, we could be sure that any effect of the preceding utterance could not be an artifact of overall developmental stage (e.g., if the younger Abe overregularized more, or favored overregularization-prone verbs, and coincidentally had more of a tendency to copy the verb type of the preceding adult utterance). Indeed, Abe's overregularizations were significantly more likely than to have been preceded by an adult stem form than correct irregulars of those verbs: 10.4% versus 4.4%, $t(90) = 1.71$, $p < .05$ one-tailed. This confirms that experimental elicitation of irregular past tense forms using the stem as the prompt will systematically overestimate children's spontaneous rate of overregularization.

4.8. Previous Estimates of Overregularization from Judgment and Correction Experiments

Another source of data that might be thought to show that children are indifferent to the past tense forms of their irregular verbs comes from Kuczaj's (1978) judgment task. In one experiment 3-9 year old children had to judge whether any member member of a group of puppets "said something silly." One puppet produced a sentence with a correct irregular past tense, a second produced an overregularization, and for verbs other than no-changers, a third produced a past+*ed* form. In a second experiment, children of the same ages produced past tenses for irregular verbs supplied in the future tense (discussed in the preceding section), then judged a puppet's version of the verb (always different from the child's version), and then judged a second puppet's version (the third possible kind of past tense form). Finally, children were offered a forced choice among the three versions, and asked which of the three they thought their mother would use. In many conditions, overregularizations were judged as acceptable at high rates, as high as 89% for stem overregularizations for the youngest children in the first experiment.

---

[2]Furthermore, providing the stem may have actively suppressed retrieval of the irregular version. Presenting an adult subject with a subset of a category of remembered words can impede retrieval of the rest (Slamecka, 1969). We thank Endel Tulving for pointing this out to us.

However here too the data are not comparable to overregularization rates from spontaneous speech. Grammaticality judgment is a signal detection task, and it is fallacious to assume that every time a child accepts or fails to correct a given form, the child's grammar deems it well-formed. Rather, just as with all yes-no data, the perceived payoffs for hits, misses, false alarms, and correct rejections affect rates of saying "yes." For children in an experimental setting this could involve a variety of demand characteristics such as the perceived politeness of rejecting or correcting another creature's language more than a given proportion of the time. In the language of the Signal Detection Theory, this defines a "criterion" or bias for saying "yes" that is superimposed on their "sensitivity" in internally representing grammatical and ungrammatical utterances as different, which we can assume is a probabilistic process. Lacking direct manipulations of bias, the best one can do in determining whether children have knowledge of irregular pasts is to compare their "yes" rate for correct irregulars versus incorrect overregularizations; if the former are higher, children must be discriminating between them correctly (see Grimshaw and Rosen, 1990). Kuczaj's data provide 15 opportunities to make such comparisons: three age groups in Experiment 1, each tested with no-change and other irregulars, and three age groups in Experiment 2, judging stem overregularizations, independently judging past overregularizations, and choosing among the three for their favorite. (Their choice of their mother's favorite form was almost identical to their own choice, and hence is not an independent data set.) Of these 15 comparisons, only 1 involved a failure to discriminate irregulars from overregularizations: the middle age group (5-7) in Experiment 2 preferred past overregularizations over irregulars or stem overregularizations in the forced choice task (though they did not even produce many such forms in the elicited production task, as noted above). In other words, the judgment data confirm that children systematically know that irregular pasts are the preferred past tense form for irregular verbs (see also Lachter & Bever, 1988).

More recently, Cox (1989) told children that a puppet "was learning to talk but was having trouble with some of his words," and the child was asked to help him say the correct words. Twelve sentences, each with an overregularized noun or verb, were provided. Children were not asked to judge the sentences, and there were no correct irregulars among the experimental stimuli, so we cannot assess children's discrimination abilities from the data. Correction performance was surprisingly poor: none of the 6 sentences with verbs was corrected by more than 16% of the children around the age of 5, and none of the 6 sentences with nouns was corrected by more than 28%, except, inexplicably, *tooths*. Since, as Cox herself notes, the children who failed to correct an overregularization did not necessarily use it themselves, a response bias against correcting the puppet too often is a likelier explanation than an absence of knowledge, especially since she told the children that the puppet was having trouble only with "some" of his words, but presented no sentences that were actually correct.

Finally, the consistent finding that overregularization rates are low and that irregulars are preferred to them helps to explain the otherwise paradoxical finding that children who have been observed to overregularize will vehemently correct their parents when they mimic the children's errors (Slobin, 1978; Bever, 1975; see Lachter and Bever, 1988). Similarly, Ervin and Miller (1963) noted that their subjects often corrected their own overregularizations; we do not know of any reports of children changing a correct irregular utterance to an overregularization.

5. Is there a U-Shaped Developmental Sequence?

What does it mean to say that children's performance is U-shaped? The textbook depiction of overregularization, as adopted by Rumelhart and McClelland (1986), assumes early correct performance on irregular pasts, followed by a preference for the regularized form and a lengthy period of indifference before correct performance is resumed. The previous section showed that one component of the supposed "U" -- a trough representing a stage in which overregularizations predominate, or even randomly alternate with correct forms -- is not true of children. However there is another aspect to the claim, namely, that a period of extremely accurate performance precedes the first overregularizations; in other words, at some point in development children get worse (even if they never get very bad).[3]

This claim has not been demonstrated quantitatively. It has passed into the textbooks on the basis of statements made by Ervin and Miller (1963), Miller and Ervin (1964), and Cazden (1968), to the effect that a few irregular forms were often used correctly by the children before they started to overregularize. For 5 children in our sample, the transcripts start at an early enough age to assess whether there is a period of uniformly correct performance preceding overregularization. For all five there were correct irregulars in the months before the appearance of the first overregularization (not surprisingly, since three of them were the ones studied by Cazden).

However, the impression of a developmental change could be a sampling artifact. Since we now know that children's overregularization rate is low, their tendency to overregularize could be unchanged throughout development, but their early samples might simply be too small to contain any examples of overregularization. Imagine drawing playing cards from a deck with replacement, looking for a black king (whose frequency in the deck is approximately equal to children's overregularization rate). One might have to draw a large number of cards before the first one appears, even if the deck is complete and properly shuffled.

It is not legitimate to test for a change in rate by comparing overregularization rates before and after the first overregularization, because the post hoc nature of the dividing line will inflate the chances of obtaining a spurious difference. One stringent test can be done as follows. If the child's overregularization rate is p, then under the null hypothesis of no change in this rate over time, the chances of the first irregular verb form in the sample being correct is $1-p$. If the likelihood of a child overregularizing an utterance is unaffected by whether or not the child overregularized the previous past tense utterance, as seems plausible, then the chances of both of the first two utterances being correct is $(1-p)^2$, the chances of the first three being correct is $(1-p)^3$, and so on. One can test whether there is an improbable run of consecutive correct irregular past tenses at the beginning of a child's records by calculating $(1-p)^n$, where n is the number of irregulars before the first overregularization. Table 4 shows the results. For Adam and Sarah, the two children with the most extensive samples, the probability that the early overregularization rate is the same as that for the entire corpus is very small ($p < .0001$). (Indeed

[3]Note that this is a qualitatively different pattern than the ones that Plunkett and Marchman (in press, 1990) try to simulate in their connectionist models. In Plunkett and Marchman (in press), the learning curves all start out at levels of performance far *less* than 100%, and then increase; the authors misleadingly term the small downward wiggles in this overall increasing curve as "U-shaped development." In Plunkett and Marchman (1990), any verb that is used once correctly and then once incorrectly is characterized as undergoing U-shaped development, which is misleading for a different reason: any stationary stochastic process (e.g., a string of coin flips) will produce local sequences with such patterns.

the probabilities would be very small even if one assumed fairly large violations of independence, such as consecutive pairs or even triples of past tense utterances being perfectly correlated in form.) For Eve, with the smallest corpus, no conclusions at all can be drawn; the first overregularization, though coming after a number of correct forms, appears exactly when it is expected to given her overall overregularization rate. For the other children, the probabilities are low, although not low enough to reject, at conventional levels of alpha risk, the null hypothesis that the overregularization rate is constant. In sum, there excellent evidence for some children that they get systematically worse as they get older; for others, the data are too sparse to tell.

Table 4
Tests of U-Shaped Development:
Correct Irregulars Preceding the First Overregularization

Child	Consecutive Correct	Overregularization	Probablity Rate
Adam	424	.018	.0006
Allison	30	.061	.15
Eve	8	.073	.54
Naomi	16	.087	.23
Sarah	253	.031	.0004

Note as well that the U-shaped sequence is a true regression in the following sense: once children begin to overregularize, they produce errors for many of the verbs they produced correctly early on (that is, the sequence does not consist of correct performance for some irregulars early, and overregularization only for newly acquired verbs, with the early correct ones eternally protected.) For Adam, 15 of his 23 overregularized types (65%) had been produced correctly at least once beforehand. For Eve and Sarah, the respective figures were 3/9 (33%) and 15 out of 26 (58%). Note that this pattern is very different from the behavior of the network model of Plunkett and Marchman (1990), whose early-acquired verbs were permanently immune from overregularization; the so-called onset of overregularization in their model pertained to its performance on newly-acquired verbs.

Are Overregularizations Preceded by Regular Past Forms? All accounts of the development of inflection must assume that the first overregularization is based on the child's attention to regular forms in the input, presumably memorized at first on a form-by-form basis, because no rule of morphology is innate. All children, then, should be capable of producing *regular* past tense forms before their first overregularization. Surprisingly, the evidence for this uncontroversial prediction is equivocal. Of the 8 children with longitudinal samples, Abe and Nathaniel's transcripts began too late to allow this test, since they overregularized in the first sample. For Adam, Eve, Sarah, and Allison, the prediction is qualitatively confirmed: Adam produced 31 regular pasts beginning 7 months before his first overregularization, Eve produced 2 beginning 3 months before, Sarah produced 12 beginning 5 months before, and Allison produced 2 beginning 6 months before. However, Naomi's and April's first overregularizations came *before* their first correct regular past tense forms, though only by a few days. Similarly, Ervin (1964) noted that for some of her subjects, there were some overregularizations before the first regular past form. These differences are all the more striking given the low overall rate of overregularization.

In the absence of suitable data sets we must leave this question unresolved, though there would be interesting implications if it turns out to be true that regulars and overregularizations can first appear simultaneously. One possibility is that a production deficit prevents children from uttering the syllable-final sequences that would allow us to credit them with controlling regular past tense forms in English; conservative learning of regulars before overregularizations could still be going on surreptitiously. The more interesting possibility, first proposed by Ervin (1964), is that there is a lag in the child's ability to perceive the regular ending in parental input, but that as soon as it is noticed, the child can extract and apply the regular rule from a small amount of evidence.

6. Factors Affecting the Temporal Course of Development

6.1. The Vocabulary Balance Hypothesis

As mentioned in the Introduction, the Rumelhart-McClelland model challenges the traditional account of overregularization, which depended on separate rote and rule mechanisms, in favor of a single mechanism that begins to overregularize because of an influx of newly-acquired regular verbs, a presumed consequence of a vocabulary growth spurt. Let us call this explanation of the cause of overregularization the Vocabulary Balance hypothesis; it is also a feature of the more recent network simulations by Plunkett and Marchman (1990). Pinker and Prince's (1988) challenge to the hypothesis rested on a demonstration that the proportion of regular verb types in longitudinal samples from Adam, Eve, and Sarah did not increase from Brown's Stages I to Stage V, but stayed at around 50%. At first this seems paradoxical: if there are only 180 irregular verbs and thousands of regulars, isn't an increase in the percentage of regular verbs a mathematical certainty after the 180th irregular is acquired, and a statistical near-certainty well before that? The answer is that Pinker and Prince's type estimates were from fairly small samples (about 700 utterances per child per stage) and hence were not pure estimates of type frequency, but something combining type and token frequency: Types with higher token frequency were more likely to have been sampled. Because the token frequency of irregulars is much higher than that of most of the regulars, it is possible that when children learn lower frequency regular verbs, they may not displace the earlier acquired irregulars. *Permit, understand, remember, misbehave,* and so on may compete among themselves for air time in children's speech, leaving general-duty verbs like *come, go, take, put, eat,* and so on to occupy a constant proportion of verb slots in conversation throughout development.

This raises an important question: should one measure types, tokens, or some other index to test the Vocabulary Balance hypothesis? Two issues are relevant: what is the psychological event that corresponds to an episode of network learning, according to Rumelhart and McClelland's theory? And what kind of changes in the schedule of learning episodes cause overregularization in pattern associator networks?

6.1.1. What is a learning episode?

Rumelhart and McClelland make the following assumptions about the real world events that correspond to a learning episode:

> The [simulation] run was intended to capture approximately the experience with past tenses of a young child picking up English from everyday conversation. Our conception of the nature of this experience is simply that the child learns first about the present and past tenses of the highest frequency verbs; later on,

learning occurs for a much larger ensemble of verbs, including a much larger proportion of regular forms.

> Although the child would be hearing present and past tenses of all kinds of verbs throughout development, we assume that he or she is only able to learn past tenses for verbs already mastered fairly well in the present tense. This is because the real learning environment does not, in fact, present the child with past-tense/past-tense pairs. Rather, it presents the child with past-tense words in sentences occurring in real-world context. The child would therefore have to generate the appropriate present tense form internally with the aid of the entire sentence and context, and this, we suppose, requires that the child already know the present tense of the word. (Rumelhart and McClelland, 1987, p.222).

The assumption here is that an episode of learning consists of hearing a past tense form, using the context to recover its corresponding stem from the mental lexicon (Rumelhart and McClelland refer to the input form as the "present" but the present tense form would include an irrelevant *-s* affix for the third person singular so "stem" is more appropriate), feeding the stem into the internal pattern associator, comparing the output with the past tense form actually heard, and adjusting the weights in response to discrepancies. A stem/past pair would be fed into the model only when an adult used the past, and the child possessed the stem in his or her vocabulary (and knew it was related to the past form). This means that the proportion of regulars fed into the past tense would be determined by the proportion of occasions that the parent used a regular past tense that the child already possessed in stem form. However, it is impossible to tell from transcripts exactly when this conjunction of parent's use and child's knowledge occurs. Instead, there are three ways to estimate the relevant proportions indirectly, each with different assumptions. For all three, it is useful to assume that all verbs have an approximately constant distribution of uses in different tenses, so we can collapse across tenses and increase sample sizes.[4]

First, if children produce regular and irregular verbs in approximately the same proportions that they process regular and irregular past tense forms in their parents' speech (which simply assumes that if a child uses a verb, he or she knows it, and that children in conversation with parents will use different verbs in roughly the same proportion as their parents), then *the proportion of verb tokens that are regular in the child's speech* indirectly estimates the proportion of regular learning episodes.

Second, since the occurrence of a parental token is necessary for a learning episode to take place, if we assume that children know a constant proportion of the verb tokens their parents address to them we can measure *the proportion of verb tokens that are regular in the parent's speech.*

In practice, Rumelhart and McClelland ignored token frequency entirely in assembling the training set for their model: every verb was fed in the same number of times, once per epoch. This assumes that a third measure, the proportion of types that are regular is the relevant factor -- though it is inconsistent with their psychological interpretation of a learning episode, which would be driven by parental tokens. Rather, the teaching schedule they modeled is more consistent with some kind of off-line learning, fed by a preprocessor: The child takes a pass through his entire verb lexicon, feeding each stem/past pair into the pattern associator once per scan. If we entertain this interpretation of a learning episode, which corresponds to Rumelhart and McClelland's actual learning schedule, rather than the token-driven interpretation they discuss, we can test the Vocabulary Balance hypothesis by trying to estimate the

[4]In fact, we have found that irregular verbs take up an even larger proportion of past tense tokens than of total verb tokens, but this difference only strengthens the conclusions we will be making on the basis of all verb tokens.

proportion of verb types in the child's vocabulary that are regular. It is important to note, however, that the meaning of this token-independent estimate is far more obscure than that of the more plausible token-driven, on-line learning scenario. Even if it turns out that the proportion of parental tokens that are regular is constant, we cannot assume that the proportion of regular learning episodes is determined by the proportion of the child's vocabulary that is regular. That is because this larger regular vocabulary could correspond to a larger number of regular types that the parent is cycling through a constant number of regular tokens in his or her speech, leaving the proportion of regular learning episodes constant.

6.1.2. How do changes in learning episodes lead to overregularization?

Assuming we know what a learning episode is, what kind of changes in the distribution of learning episodes lead to overregularization, according to the Vocabulary Balance hypothesis?

First, it is clear that both type and token frequency have important consequences. The Rumelhart-McClelland model overregularized because it changed its connection strengths with each input pair in a direction that reduces the discrepancy between computed and input past forms. After the first epoch in which the model was suddenly bombarded with regulars, about 80% of the changes the model made were designed to make it more likely to generate regular forms, because 80% of the inputs were new regulars. Many of the changed connections involved links from phonological features that were also shared with irregulars (since most irregulars are phonologically similar in some way to at least some regulars). Because the network did not have enough specific feature units to register each verb on its own set of units, the overlap was high enough that each irregular was represented by many units whose links had just been adjusted to help produce the regular ending, and overregularization resulted. This effect would obviously depend strongly on the number of regular types, because the wider the range of regular forms that are fed in, the greater the probability that a given phonological feature of an irregular verb will be shared by some regular verb and hence develop stronger links to the incorrect regular pattern. But this effect can also be mitigated by token frequency: if, say, each irregular had been repeated 4 times for each regular (reflecting the real-world higher token frequency of irregulars), the links that joined features unique to the irregulars to their corresponding irregular past forms would have been strengthened several times during the epoch to reduce the errors with such forms, and overregularization would be less likely. Indeed since one of the noteworthy properties of the Rumelhart-McClelland model is its distributed phonological representation of words, with no units dedicated to words per se (see Pinker & Prince, 1988), there is no physical basis for a distinction between types and tokens in the model; only feature-to-feature mappings, whether they be from a single word or a set of similar words, are represented. The actual behavior of the model will depend on the number of regular types, the phonological range of the regular types, their degree of overlap with irregulars, the token frequencies of both irregulars and regulars, and other factors. In any case it is clear that all things being equal overregularization of an irregular does depend on the ratio of regular to irregular tokens, and hence is relevant to testing the Vocabulary Balance hypothesis.

Second, the percentage of regular learning episodes at a given time is not the relevant factor in predicting overregularization. In the Rumelhart-McClelland model, unlike children, the process of recovery from overregularization begins immediately after its onset (see Figure X), correct irregulars

predominate within a few epochs, and at asymptote they are produced most of the time, all with a constant level of 80% regular learning episodes. This is an obvious property of any model that is designed to perform correctly at asymptote: even with the most unfavorable proportion of regular episodes, the irregulars must eventually reassert themselves. Overregularization is a short-term consequence of the *increase* in the percentage of regular episodes with development: though a model could learn to overcome any particular level of dominance of regulars, this adjustment cannot take place instantaneously, and influxes of regulars will cause temporary overregularization, before the crucial links between nodes unique to an irregular and its idiosyncratic past have been sufficiently strengthened. For this reason, the difference between the rapid recovery of the Rumelhart-McClelland model and the protracted period of overregularization of children does not speak against the model. It is possible that as children learn more and more words, new regulars are constantly washing over them; no sooner do they adjust their irregulars to the leveling effect of one wave of regulars than a new wave comes in. Thus the proper test of the Regular Vocabulary Balance hypothesis involves a correlation between the most recent *increase* in regular learning episodes and the current rate of overregularization.[5]

Third, for the analyses in which types are being examined, it may not be the proportion of verb types that are regular that is the relevant predictor. The problem is that the competition in pattern associators is not between the regular pattern and a single irregular pattern shared among all the irregulars. Rather, irregulars are different from each other, not just from the regulars. Imagine that at one stage, there are six different irregulars, each with a different change (e.g., *go/went, come/came, hit/hit,* etc.) and six regulars. At the next stage, there are twelve different irregulars, each with a different change, and twelve regulars. The proportion of regulars in the sample remains the same, but the ratio of regulars to *any particular irregular pattern* has doubled. Therefore in this scenario it is the number of regulars, not the proportion of regulars, that would predict overregularization. The scenario is not accurate, however: the ratio of learning episodes for the regular pattern to a given irregular pattern would be identical to the number of regular only if every irregular were totally idiosyncratic. But virtually all the irregulars share their patterns of change with other irregulars, so the calculation is too extreme. Consider a scenario that is extreme in the other direction: the six irregulars in phase one fall into three classes (e.g., *sing/sang, ring/rang, feed/fed, breed/bred, wear/wore, tear/tore*), and the new irregulars in phase two fall into the same classes (e.g., *spring/sprang, lead/led, swear/swore*). Here the ratio of regulars to any irregular vowel change pattern is 3:1 in both phases, and we would expect overregularization to be less likely; only a change in the proportion of regulars would clearly induce it. In reality the situation is likely to be somewhere between these extremes because while English irregulars do fall into a restricted number of kinds of change, we would expect the number of patterns in a child's vocabulary, not just the number of irregulars per pattern, to increase somewhat with development. Therefore it is not clear whether overregularization rates should be correlated with the proportion of total types that are regular (appropriate if all new irregulars fall into old patterns and hence protect old irregulars), or the number of types that are regular (appropriate if each new irregular is unique), and we will examine both.

---

[5]In recent experiments on the behavior of connectionist networks at learning inflectional mappings in sets of artificial verbs with different training schedules and vocabulary mixtures, Plunkett and Marchman (1990, in press) have confirmed that both token frequencies and rate of vocabulary increase have direct effects on the tendency of standard connectionist models to produce outputs analogous to overregularizations.

6.1.3. Summary of Tests

Any of the vocabulary measures listed above could increase with development; indeed the proportion of regular types in the child's vocabulary *must* increase by mathematical necessity. But any global increase in some factor with age (e.g., the child's height) is confounded with dozens of other factors, so any correlation over months with overregularization rate does not imply that it is responsible for overregularization rate. At the very least one must correlate *changes* in the factor with the phenomenon of interest. This can be done in two ways.

We will compare the monthly rate of increase in each vocabulary factor for the months before the first overregularization with the monthly rate of increase for the months during which overregularization is taking place (i.e., the first month containing an overregularization and all the months after it.) The dividing line is not perfectly clear for Sarah. Brown (n.d., see Pinker & Prince, 1988) suggested that her first potential overregularization, *heared* at 2;10, could have been a mispronunciation of *heard*; the first unambiguous overregularization does not occur for another five months. However, this particular pattern of distortion is not independently motivated by Sarah's other odd pronunciations at that time, and excluding it has the effect of weakening the evidence for the Vocabulary Balance hypothesis; therefore we count it as her first overregularization.

A more precise estimate comes from correlating the monthly rate of increase in a vocabulary measure between month time t and month $t + 1$ with the overregularization rate at time $t + 1$. This analysis combines the factors that differentiate the pre-overregularization stage from the overregularization stage and and the factors that cause overregularization to be more frequent in one month than another during the overregularization stage. According to the Rumelhart-McClelland model, there is no qualitative difference between the two. Both of these comparisons will be presented for the parent's tokens, the child's tokens, and the child's types.

6.2. Parental Tokens

The proportions of adult verb tokens that are regular are plotted for Adam, Eve, and Sarah in Figures 14, 15, and 16. The figures include not only the child's parents but the other adults speaking to him or her in the transcripts. For all three children, only about a quarter of the parental verb tokens were regular, and this did not change over the course of development. As mentioned, this is a consequence of the fact that most of the high-frequency verbs that are indispensable for casual conversation are irregular, and do not move aside to make way for the more numerous but lower-frequency regulars.

The proportions are similar before and after overregularization begins: For Adam, 26% before, 26% during; for Eve, 26% before, 22% during; for Sarah, 23% before, 23% during. More importantly, the rate of change in the proportions was not systematically larger during the overregularization period: For Adam, +1.1 percentage point per month before, -0.1 during; for Eve, +1.1 versus -0.6; for Sarah, 0.0 versus +0.2. The correlations between the rate of monthly change in proportion regular and the child's overregularization rate are in the wrong direction for Adam (-.43) and Eve (-.35) and close to zero (.04) for Sarah.

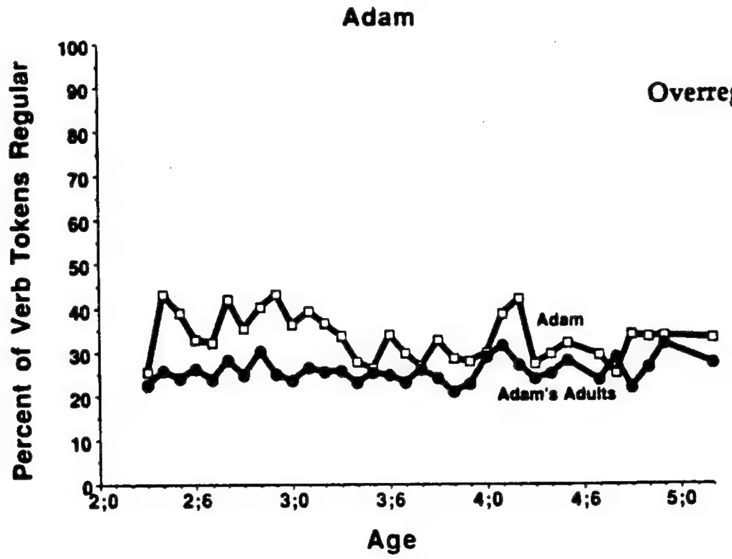


Figure 14

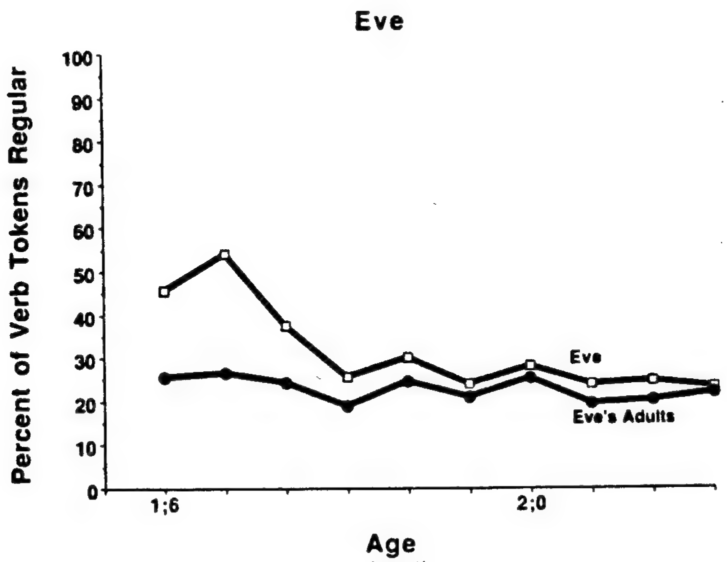


Figure 15

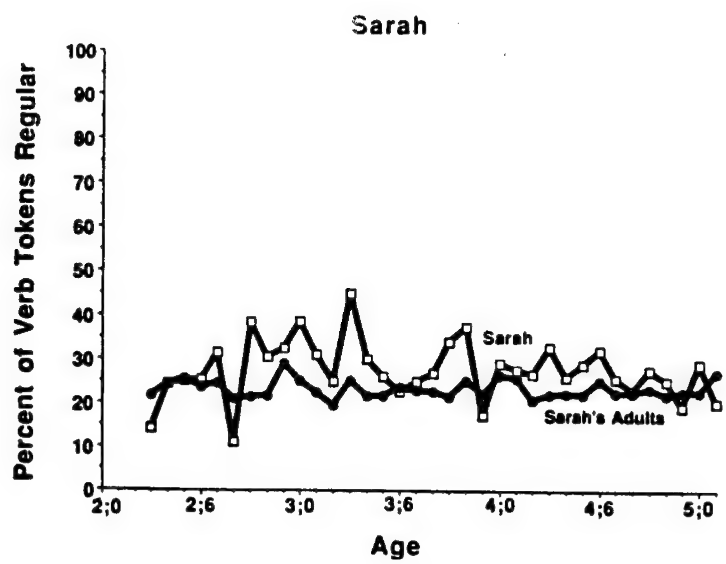


Figure 16

Figures 14-16. Percentage of verb tokens that are regular for Adam, Eve, Sarah, and the adults conversing with them.

6.3. Child's Tokens

The proportions of child's tokens that are regular are plotted on the same axes as their adults' proportions, also in Figures 14-16. The proportion oscillates between 25% and 43% for Adam, actually declines for Eve from around as high as 53% (possibly sampling error) to a steady state between 25% and 30%, and shows some early dips to the teens for Sarah before oscillating within the 20% - 40% range. The reasons for regular verbs being used a steady minority of the time are no doubt the same as for the adults.

The proportion of tokens that are regular does not systematically increase after overregularization begins: 36% versus 32% for Adam; 49% versus 27% for Eve; 24% versus 28% for Sarah. The monthly rates of change for these figures are actually higher before overregularization than during: +2.1 percentage points versus -0.3 percentage points for Adam; +8.3 versus -3.8 for Eve; +4.0 versus -0.6 for Sarah. The correlations between monthly change in percent tokens regular and the child's overregularization rate are .01 for Adam, -.28 for Eve, -.09 for Sarah.

6.4. Child's Types

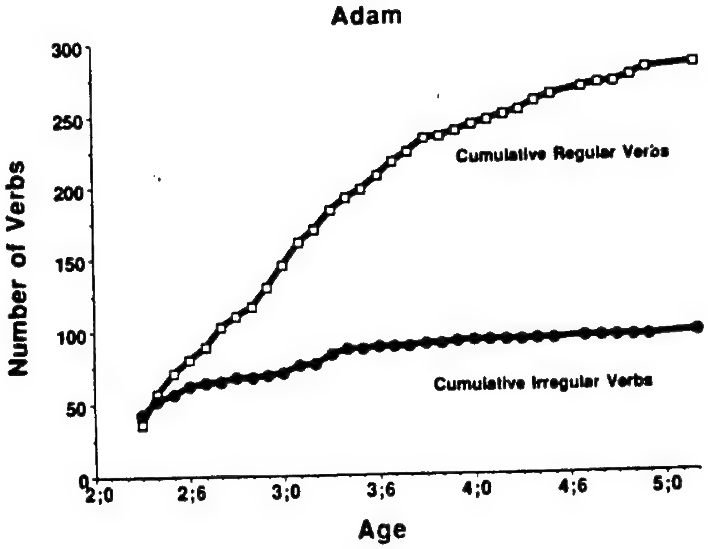
As mentioned, in one sense this is the least psychologically realistic measure to focus on, because it assumes that rule learning is an off-line pass through the child's vocabulary, and as such does not correspond to Rumelhart and McClelland's psychological assumptions. But it is worth focusing on both because it is the form of the Vocabulary Balance hypothesis that is most likely to be consistent with some developmental trend (since the proportion of regular types must increase with development) and because it literally corresponds to the training sequence given to the Rumelhart-McClelland model.

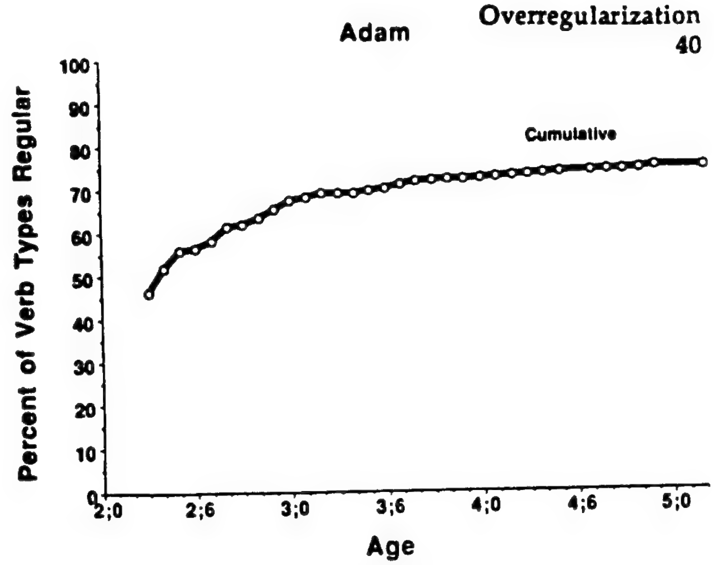
Measuring the proportion of children's vocabulary that is regular is an extremely difficult problem, for it inherits all the notorious pitfalls in estimating children's vocabulary size in general (see, e.g., Seashore & Eckerson, 1940; Templin, 1957; Lorge & Chall, 1962; Miller, 1977; Moe, et al., 1982, for extensive discussion.) The source of the problems is that we are confined to the actual words that children used in samples. Obviously the child will only use a small fraction of his or her total vocabulary in any given sample. Since high-frequency verbs are more likely to appear than low-frequency verbs, the number of low-frequency verbs will be systematically underestimated. And since there are more low-frequency regulars than low-frequency irregulars, counting types per sample will systematically underestimate the proportion of regulars (this was a problem with Pinker and Prince's 1988 estimates). There is no completely adequate solution to the problem of measuring children's vocabulary, but there are various estimates that can be examined, and at the very least the direction of changes in the proportion that is regular can be compared for months associated with different levels of overregularization, and the resulting conclusions compared across different methods in an attempt to arrive at converging conclusions.

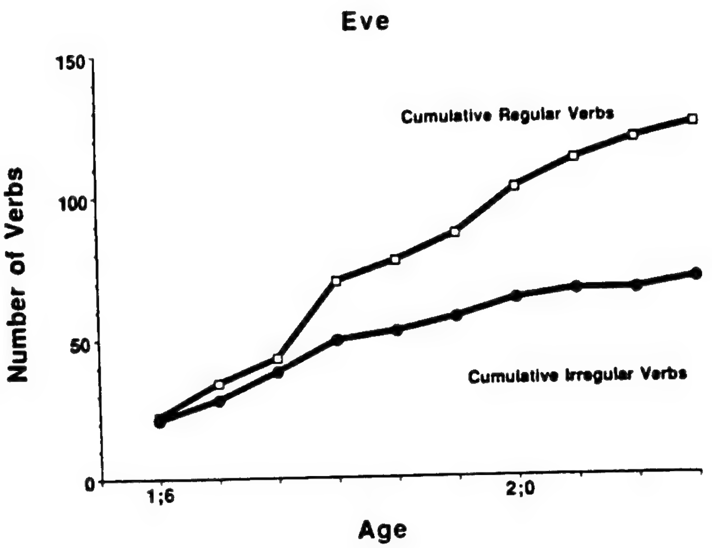
6.4.1. Method 1: Cumulative Vocabulary

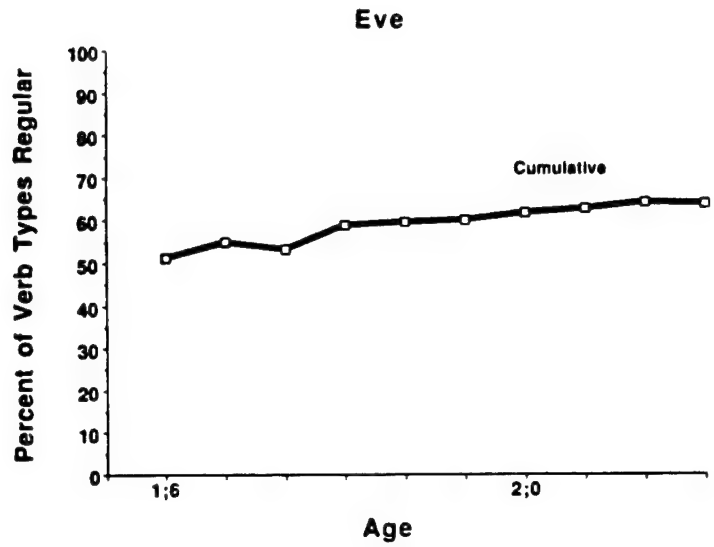
One measure that is designed to be generous to low-frequency forms is the child's *cumulative* vocabulary totals. That is, one assumes that the child never forgets; if a word is used in a given month, it is credited to the child's vocabulary from then on. Figures 17-19 shows the children's cumulative vocabulary growth for regular and irregular verbs; Figures 20-22 show the proportion of cumulative vocabulary that is regular. Table 5 shows these figures for Adam, Eve, and Sarah at the month before

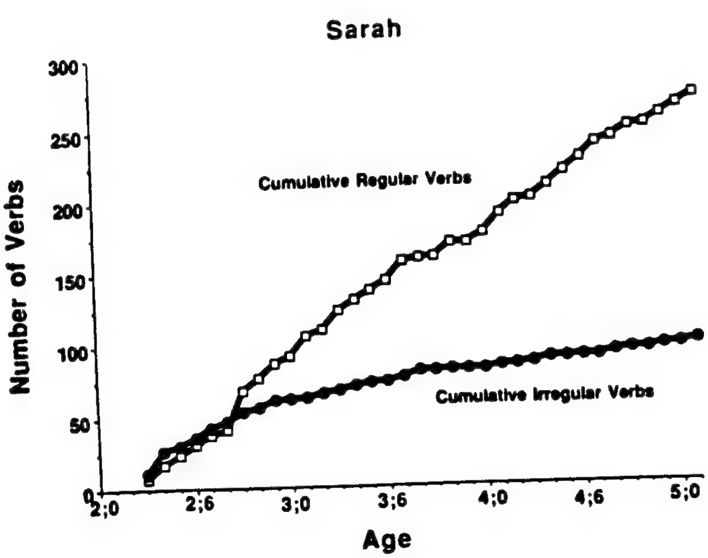
overregularization begins and at the end of their transcripts. By mathematical necessity each child possesses a larger cumulative regular vocabulary later in development than earlier, and as expected, the regulars take up a larger proportion of the child's total verb vocabulary later. However, as the decelerating vocabulary curves (most visible for Adam) suggest, for both the number of regular verbs and for the proportion of verb vocabulary that is regular, the rates of increase are much larger for the stages before than during overregularization, contrary to the Vocabulary Balance hypothesis. Similarly, the monthly change in number of regular verbs and proportion of verbs that are regular correlates negatively with overregularization rate for all three children. These rates of change are also summarized in Table 5.











Figures 17-22. Cumulative regular and irregular verb vocabulary for Adam, Eve, and Sarah, and the proportions of verbs in their cumulative vocabularies that are regular.

Table 5
Growth of Cumulative Regular and Irregular Vocabulary

	Adam	Eve	Sarah
Average Monthly Rate of Increase Before First Overregularization			
Number of Regular Verbs	11.6	12.0	10.2
Proportion of Verbs Regular	2.4	3.6	2.9
Month Before First Overregularization			
Number of Regular Verbs	117	34	68
Number of Irregular Verbs	68	28	53
Total	185	62	121
Percent Regular	63	55	56
Average Monthly Growth During Overregularization			
Number of Regular Verbs	6.6	11.3	7.2
Proportion of Verbs Regular	0.5	1.1	0.6
At End of Transcripts			
Number of Regular Verbs	283	124	269
Number of Irregular Verbs	99	70	99
Total	382	194	368
Percent Regular	74	64	73

Unfortunately the decelerating vocabulary curve and concomitant negative correlation with overregularization may be a direct consequence, even an artifact, of the cumulative measure for vocabulary. Cumulative vocabulary is equivalent to sampling without replacement. Imagine that parents use 500 regular verbs, with equal token frequencies, when speaking to their children, at all ages. Imagine that every month children attend to and acquire 10% of the verbs they hear, and produce every word they have acquired at least once. At the end of Month 1 their cumulative vocabulary is 50 words. At the end of Month 2 it is not 100 words but only 95 – the 50 they learned in Month 1, plus 10% of the 450 words that they had not previously acquired in Month 1 (i.e., 45 words; the other 5 they attended to don't count because they had already been acquired.) In Month 3 their cumulative vocabulary will be 136, reflecting the addition of only 40 new words (10% of the 405 remaining), and in Month 4, 172 (36 new words). In other words, new words will be acquired at a faster clip early in development than later, even with a constant learning rate and constant number of words in the environment. If there is indeed a relatively limited set of regular verbs for the child to acquire during the preschool years, this is one reason why the Vocabulary Balance hypothesis could be false.

However these sampling considerations could also mean that the decelerating growth curve is an

artifact. Imagine that the child possesses a constant 500 words throughout development, but only manages to produce 50 during a month's worth of samples. Following the same arithmetic as described above, the fact that cumulative vocabulary is a form of sampling without replacement means that we as investigators will spuriously credit the child with having "acquired" fewer and fewer new words with each succeeding month. It is very difficult to tell to what extent the curves in Figures 17-19 represent a genuine sampling effect in vocabulary learning, a sampling artifact in measuring vocabulary from production data, or both. Thus it is important to supplement these direct estimates with some indirect measure that is free of this possible bias.

6.4.2. Method 2: Jackknife Estimates of Capture-Recapture Patterns

There is a family of techniques commonly used in biology and demography for estimating populations sizes from multiple samples. The simplest version is commonly known as *capture-recapture* (see Seber, 1986). For example, here is an idealized example of how one might estimate the number of squirrels in a forest. Trap 50 squirrels, paint their tails orange, release them, allow enough time for them to diffuse through the forest, trap 50 squirrels again, and see how many have orange tails. If there are 10 such recaptures, then the first trapping session must have represented 10/50 or 1/5 of the total population. Since 50 were trapped initially, the forest population must be 250.

This can be applied to vocabulary estimation as follows. A verb is a squirrel, a transcript is a trapping session, and a verb that appears in two successive transcripts has been recaptured. The proportion of verbs at t_2 that also appeared at t_1, multiplied by the number of verb types recorded at t_1, is an estimate of the vocabulary size. Note that this procedure avoids the possibly artifactual deceleration in vocabulary acquisition suggested by cumulative measures. If a child had a static vocabulary of 500 words, 50 of which were recorded in each sample, then the second sample would consist of 5 of the words that appeared in the first sample (1/10 x 50) and 45 new words (1/10 x 450), and the recapture rate of 1/10 (5/50), multiplied by the first sample size (50), would yield the correct figure of 500, and this would be true of every pair of successive samples. Note as well that if new verbs are acquired between t_1 and t_2, the estimate will be an unbiased estimate of the vocabulary at t_2. Imagine an idealized case in which 100 baby squirrels were born between capture and recapture. Orange-tailed squirrels are now recaptured with probability 50/350, so the recapture proportion is 1/7. Seven times the capture sample size of 50 is 350, the true population size at recapture.

Unfortunately, unequal token frequencies for different verbs lead to systematic underestimates if one were to use this procedure unmodified. Imagine that some squirrels are more trap-shy than others – in the simple case, 40% of the squirrels might be "shy," where "shyness" means that the probability of blundering into a trap (1/5, in our example) is cut in half. One expects the the first sample to capture 30 of the 150 bold squirrels, but only 10 of the 100 shy squirrels. Sixteen percent (40/250) of the squirrels in the forest now have orange tails, but the second trapping session will only recapture 7 of them: 6 bold ones (1/5 x 30) and 1 shy one (1/10 x 10). Since the second sample consists of 40 squirrels in all ((1/5 x 150) + (1/10 x 100)), the 7/40 recapture rate, multiplied by the 40 squirrels in the first sample, yields an estimate of 229, 21 less than the true figure. For children's vocabulary, verbs with lower token frequencies are "trap-shy," and because many of them are regular, one would obtain systematic underestimates of total verb vocabulary, regular verb vocabulary, the percent of vocabulary that is

regular, and the rate of increase in regular vocabulary (though comparisons of higher- versus lower-growth months might still be roughly accurate).

Biostatisticians have dealt with the trap-shyness problem by applying the "generalized Jackknife" estimator to the capture-recapture methodology, first developed by Burnham and Overton (1978, 1979), and extensively investigated by Otis, Burnham, White, and Anderson (1978). Instead of two trapping sessions, there are k of them. The numbers of squirrels that have been captured only once, twice, three times, and so on, are tallied. The effects of unequal capturability can thus be estimated by taking into account the distribution of multiple recaptures. While the simple capture-recapture estimate assumes a uniform distribution of capturability, with the extra information of number of recaptures one can assume that individual capture probability is a random variable from an arbitrary distribution. Otis, et al. (1978) found empirically that the Jackknife estimator produces accurate estimates if many individuals are caught a relatively large number of times - that is the multiple recapture rate across multiple samples is high.

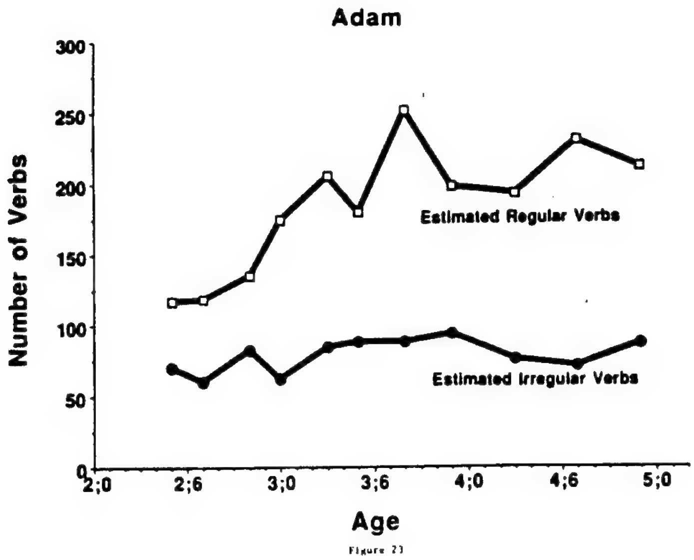
In our case this consists of comparing sets of five consecutive transcripts, and counting how many verbs were used in each of the 5 transcripts, how many in only 4, and so on. (To make the estimates for the three children comparable, for Sarah we use 5 consecutive *pairs* of transcripts because her speech was sampled for an hour once a week whereas Adam's and Eve's was sampled for two hours once every two weeks.) As required, many verbs in fact appeared in multiple transcripts from such sets. This set of numbers is fed into the Jackknife algorithm, providing an estimate of vocabulary size for that period. With non-overlapping sets of 5 consecutive transcripts, we obtain independent estimates for different ages. The deceleration in cumulative estimates is eliminated, as is the underestimation inherent in simple capture-recapture estimates.

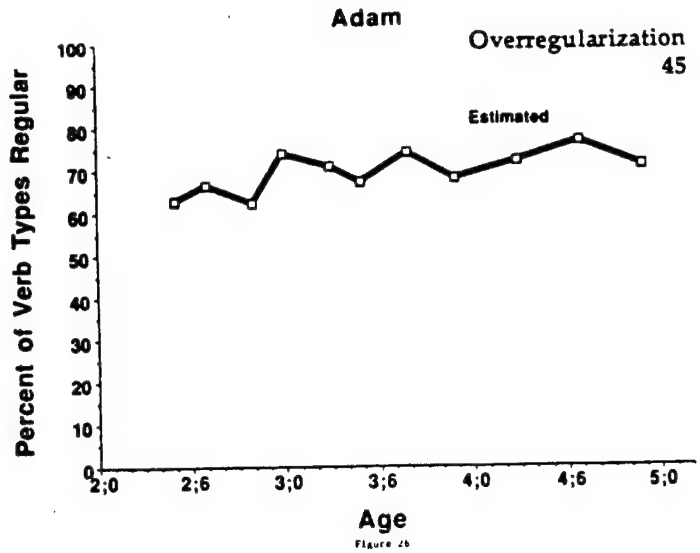
The estimates span periods of 2 1/2 months rather than a single month, which has both disadvantages and advantages. The growth estimates are temporally coarser, and there is no vocabulary estimate at all for Eve restricted to the period before her first overregularization. However a larger temporal window may catch effects of vocabulary growth that act over longer time spans than the one-month window used so far.

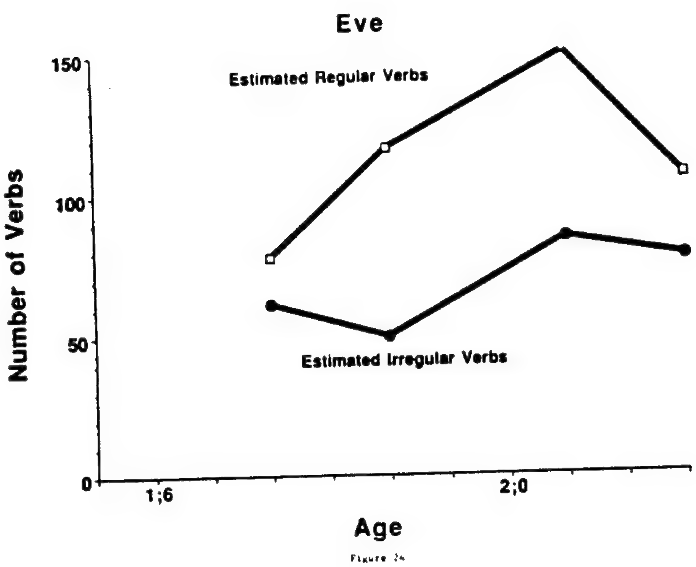
The estimator is not free of complications. Because the kind of context in which the recording takes place is similar in all recording sessions, those verbs most appropriate to those contexts will be recorded more often. (In the ecological literature, it has also been noted that achieving equal capturability is impossible, even with randomized capture locations on each sampling occasion; see Chao, 1987.) Furthermore there is a free parameter that must be decided upon in calculating the estimates: the "order" of the estimate, corresponding to the maximum number of recaptures (out of 5, in our case) that are counted in the calculations. Higher order estimates have lower bias but higher variance; there is a complex procedure for selecting the optimal order for a given estimate. For simplicity's sake, we will uniformly report estimates of order 4. We have found these generally to be the highest of the estimates of different order, especially for regular verbs, and hence they are fairest to the Vocabulary Balance hypothesis. But in any case we also found that the growth curves for different order estimates are almost perfectly parallel, so the correlations we calculate are not notably affected by this choice. With these considerations in mind we can cautiously compare periods of high regular vocabulary growth with

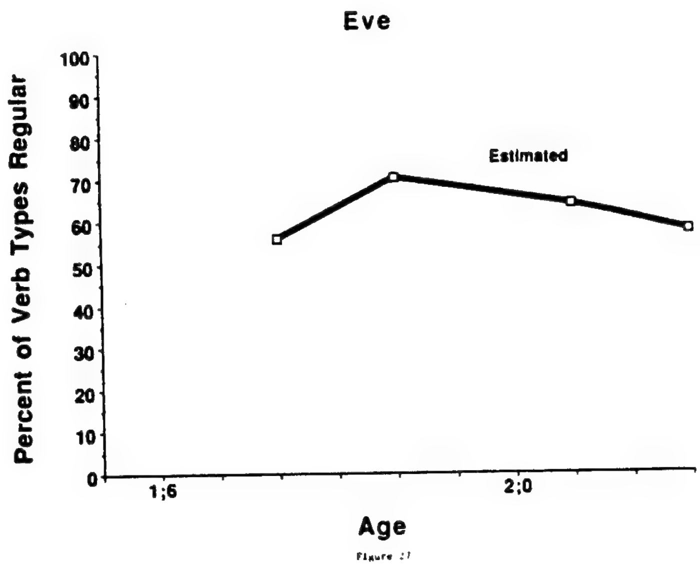
periods of low regular vocabulary growth, even if the magnitudes of particular increases and totals are not taken as perfectly accurate estimates.

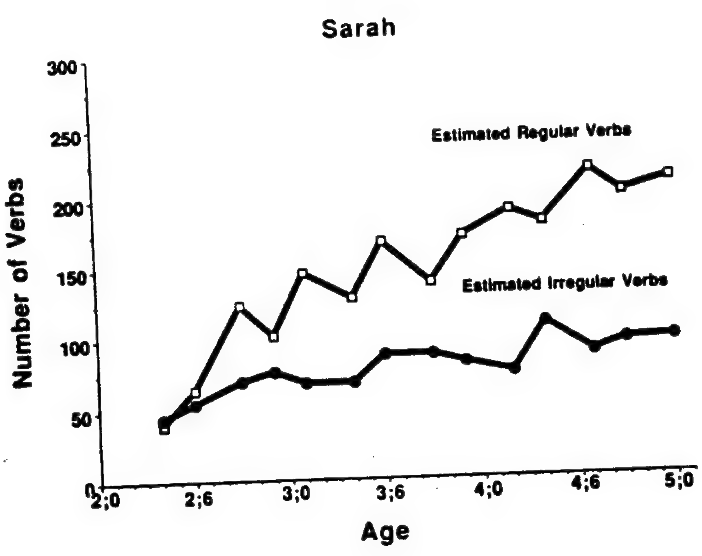
Figures 23-25 shows that the vocabulary estimates obtained from this method are higher at young ages and somewhat lower at older ages than the cumulative estimates, eliminating the the severe deceleration that was inherently unfavorable to the Regular Vocabulary Balance hypothesis. They also do not display the unlikely constant 50/50 regular-irregular ratio that would correspond to interpreting Pinker & Prince's figures as estimates of types. Table 6 shows the estimated vocabulary before the first overregularization and at the end of the transcripts, and the average rate of growth per interval during the period preceding each of these points. The number of irregular types shows a very small increase with time; the number of regulars shows a larger one. As Table 6 shows, for Adam, the rate of increase in the proportion of verbs that is regular is larger during the overregularization stage than before it (1.1 versus -0.2 percentage points per 5-sample interval); rate of increase from one interval to the next is positively (though not significantly) correlated with the overregularization rate at the end of the interval (r =.12, p > .25). These trends, however, constitute the only comparisons from among all those we have performed that are in a direction consistent with the Vocabulary Balance hypothesis. For Adam there was virtually no difference in the number of regular verbs acquired per interval during the overregularization stage than before it (9.7 versus 9.1 new regular verbs per interval), and the correlation between the size of the increase between intervals and the overregularization rate for the second interval is negative (r = -.05). Sarah acquired regular verbs at a faster rate before her overregularization stage than during it (42.2 versus 10.2 additional verbs per interval), and these increases correlated negatively with their ensuring overregularization rates (r = -.22). Similarly, the proportion of her vocabulary that was regular increased much more rapidly before than during her overregularization stage (8.3 versus 0.4 percentage points per interval), and correlated negatively (-.14) with overregularization rate. For Eve, the temporal coarseness of the Jackknife estimates prevents a before-and-after comparison of vocabulary composition, but both the increase in number of regular verbs and the increase in proportion of verb vocabulary that is regular correlate negatively with overregularization rate at the end of the relevant interval.











Figures 23-28. Jackknife-estimates of regular and irregular verb vocabulary for Adam, Eve, and Sarah, and the proportions of estimated verbs that are regular.

Table 6
Growth of Regular and Irregular Vocabulary, Jackknife Estimates

	Adam	Eve	Sarah
Average Rate of Increase Before First Overregularization			
Number of Regular Verbs	9.1	*	42.2
Proportion of Verbs Regular	-0.2	*	8.3
Month Before First Overregularization			
Number of Regular Verbs	136	*	124
Number of Irregular Verbs	82	*	69
Total 218	*	193	
Percent Regular	62	*	64
Average Rate of Growth During Overregularization			
Number of Regular Verbs	9.7	*	7.8
Proportion Regular	1.1	0.6	0.4
At End of Transcripts			
Number of Regular Verbs	213	106	209
Number of Irregular Verbs	87	78	96
Total 300	184	305	
Percent Regular	71	58	69

*Eve had too few samples before her first overregularization to yield Jackknife estimates.

With all the hazards of vocabulary estimates it would be reassuring to compare ours with previous ones in the literature. Horn (1927) amassed 489,555 tokens of the speech of kindergarten children, comprising 7,097 types. She reports the 1003 most frequently used words. Unfortunately this decision will result in a systematic underestimate of the number of lower-frequency words a given child possesses, which in turn systematically underestimates regular verbs. The list contains 97 regular verbs and 74 irregular verbs, or 57% of verb types regular. These figures are smaller than the cumulative and Jackknife proportions we find for Adam and Sarah at the end of their transcripts, but this is to be expected because of of the exclusion of lower-frequency words; when we look at single samples of Adam's and Sarah's speech, which introduces a similar bias, we get figures of up to 58%, extremely close to Horn's estimate. Moreover Horn's data for the proportion of verb tokens that are regular, which are unaffected by this sampling bias, are almost identical to our estimates: 24,581 regular tokens and 80,370 irregular tokens, or 76%.

Even better estimates come from Moe et al.'s (1982) figures for first grade children (mean age 6;9), because they report all verb tokens, not just the most frequent ones. Their lists include 418 regular verb types and 108 irregular types, corresponding to 79% of verb types being regular, and 33% of verb tokens. These figures closely match the estimates shown in Tables 5 and 6 for Adam and Sarah at the end of their transcripts: Adam (5;2): 74% cumulative types regular, 71% Jackknife-estimated types regular, 33% tokens regular; Sarah (5;1): 73% cumulative types regular; 69% Jackknife-estimated types regular; 20% tokens regular).

In sum, virtually without exception our estimates of children's types, adults' types, and children's tokens provide no quantitative support for the Vocabulary Balance hypothesis. Regular verbs remain a roughly constant proportion of adults' and childrens' conversational tokens, and never dominate. Regular types -- which in any case do not correspond to on-line learning episodes -- necessarily increase with development, both absolutely and as a proportion of total verb vocabulary, but the sizes of these increases do not systematically correlate with children's tendency to overregularize, which is what the Vocabulary Balance hypothesis requires. It appears that something endogenous to the child's grammatical system, and not a change in either the environment or the vocabulary, causes overregularization errors to begin.

6.5. Obligatory Past Tense Marking as a Predictor of Overregularization

We have found only one measure that does seem to be related to the onset of overregularization. Kuczaj (1977) summarizes a pattern in his longitudinal and cross-sectional data as follows:

> Apparently once the child has gained stable control of the regular past tense rule, he will not allow a generic verb form to express "pastness," which eliminates errors such as *go*, *eat* and *find*, but results in errors like *goed*, *eated*, and *finded*. (p. 593)

The suggestion is that the children avoid using stem forms in past tense contexts during the period they are overregularizing. Although very young children leave most verb forms unmarked, it is logically possible that as they get older, two kinds of errors with irregular verbs -- no marking and overregularization -- decrease in tandem. Instead, Kuczaj suggested that they are in a reciprocal relationship; overregularization errors appear as no-marking errors (stems in obligatory past tense contexts) disappear.

Note that this is a straightforward prediction of the traditional rote-plus-rule theory. Overt tense marking in the matrix clauses of English is *obligatory*, though children cannot be born knowing this. If the learning and reliable application of this syntactic constraint is not instantaneous, there will be a period in which children will fail to mark past tense forms at all. During that time, retrieval failure for a given irregular past tense form will not necessarily result in an overregularization: the child has the option of uttering the bare stem. But once the child realizes that tense marking is obligatory, he or she will have to mark a verb in a past tense context in one form or another. If search for a stored form for a verb turns up nothing (either because it is regular or because the irregular past tense form is recalled with less than perfect accuracy) the regular rule will step in to supply the verb in the tensed form that the syntax demands.

Hence the crucial factor in the onset of overregularization is *obligatoriness*. If the probability of

retrieving a stored past tense form given that it is searched for is, say, 4%, but the young child is only searching for past tense forms 10% of the time he should be, then observed overregularizations would only be seen 0.4% of the time, too infrequently to turn up in most samples. And if the development of the regular rule itself takes time, so that at first it is not applied with perfect reliability even when it is called, the likelihood of observing an overregularization in the early stages would be even lower. Only when at least one of these two processes (seeking a past tense form for an irregular, and successfully applying the regular rule when it is called) will overregularizations be observed.

What is the evidence for the Obligatory Tense-Marking Hypothesis? Abe began overregularizing at the beginning of the samples (Kuczaj, 1976, gives an example from 2;5) so a straightforward before-and-after comparison cannot be done, but according to Kuczaj's (1977) data, at 2;6 Abe failed to mark the past tense of irregular verbs in obligatory contexts 40.9% of the time, and his overregularization rate was 2.4%. The tables in Kuczaj (1977) report that Abe marked 100% of irregular past tense forms in the samples taken from 3;0 through 5;6, during which time he his overregularization rate was much higher, ranging from 4.2% to 45%. Regular verbs, too, were marked for past tense 100% of the time after the first month. The more detailed monthly data from Kuczaj's thesis confirm that no-marking errors were rare throughout the overregularization phase: after 2;9 only 4 out of 28 months contained a no-marking error, and only 0.4% of all irregular verbs in past tense contexts were left unmarked.[6]

Kuczaj's (1977) tables of cross-sectional data from 14 children between 2;6 and 5;6 allows us to assess the possible dependence of overregularization on the diminishment of no-marking errors for irregular verbs. Since no-marking errors decrease with age and overregularizations increase or decrease (depending on the age; for Kuczaj's subjects, they decrease), it is important to try to measure the correlation between overregularization and no-marking holding age constant. In a multiple regression with overregularization rate as the dependent variable, and age and percentage of irregular verb usages that were no-marking errors as independent variables, both age ($r = -.52$) and percentage of no-marking errors ($r = -.27$) were negatively correlated with overregularization rate, and the unconfounded effects of each, partialing out the effects of the other, were statistically significant (for age, $F(1,11) = 11.91$, $p < .01$; for no-marking percentage, $F(1,11) = 7.03$, $p < .05$; age and no-marking percentage correlated -.41).

Brown's (1973) data for Adam, Eve and Sarah are also consistent with obligatoriness being the chief predictor of the onset of overregularization. Brown attributed control of a morpheme to a child only when they met the stringent criterion of using it in 90% of its obligatory contexts in six successive hours of speech. According to Brown's Figure 14 (p. 271), Adam passed this test shortly after 2;11; his first overregularization was recorded at 2;11. Sarah also exceeded this criterion around 2;11, and her first overregularization was recorded at 2;10. At first Eve appears anomalous: she overregularized at 1;8, but

[6]We can also speculate that Abe's near-exceptionless marking of past tense is in part responsible for his unusually high overregularization rate. We know that Adam and Sarah, for example, marked irregular pasts more than 90% of the time, but we don't know how much more. If they continued to leave as many as 10% of their irregular pasts unmarked, presumably those are just the usages in which they were likely to have failed to retrieve the irregular past, and which would have been overregularizations if they, like Abe, would let no past tense usage go unmarked. Adding up to ten percentage points to their overregularization rates would narrow the gap between them and Abe in the distribution. Of course the discrepancy can alternatively be attributed to Abe's having poorer retrieval of irregular forms from memory, or to some combination of these individual differences.

failed to reach Brown's criterion for irregular pasts even at the end of her transcripts at 2;3, at which age she was almost as linguistically sophisticated as Adam and Sarah at 5. Unfortunately Brown supplied only the ages of reaching criterion, not the full learning curves, so we cannot tell whether Eve was marking pasts none of the time or 89% of the time when she began overregularizing. We went back to Eve's transcripts from the three available months before and including overregularization, redoing Brown's analysis of the proportion of time an irregular past tense form was supplied when obligatory. The rates were 100% and 75% for the two months preceding the first overregularization, and 85% during that month, which are very high proportions, missing Brown's criterion by a hair. Thus all three children began marking the past tense of irregular verbs almost all of the time around the age at which they started to overregularize.

In sum, four different data sets show a close relationship between obligatory marking of irregular past tense forms (i.e., diminution or disappearance of no-marking errors) and the onset or degree of overregularization. The results of this section, then, suggest that the crucial antecedent to overregularization is not a shift in the statistical preponderance of regular verbs in the child's input or vocabulary, but the full deployment of the syntactic principle that tense-marking is obligatory in certain contexts. An insistence on marking the past tense, combined with imperfect retrieval of irregular forms from memory, and a regular process that is capable of applying to any stem, even if closely tied to an irregular past, sets the stage for overregularization errors.

7. Lexical Factors in Overregularization

Although overregularization rates are low in general, not all verbs are overregularized at the same rate, and some verbs are overregularized by some children at some stages more often than they are produced correctly. By examining what it is about a verb that makes it more or less likely to be overregularized, one can test many hypotheses about the psychology of overregularization.

In this section, we correlate a variety of lexical factors with overregularization rates across a large set of irregular verbs. To minimize averaging artifacts, we report these correlations for individual children, focusing on Abe, the most prolific overregularizer. We also report individual correlations for the other 18 ChilDES children with individual transcripts that overregularized at least once. When we report averages for these correlations, and tests of the average against a null hypothesis of zero, the correlation coefficients (r) are first transformed to Fisher's z; means are obtained by averaging the z-scores and transforming the mean z back to a correlation coefficient. In addition, we computed a single measure that aggregates overregularization tendencies for verbs across these 18 children. In order to prevent such a measure from being artifactually influenced by individual children with extreme values and idiosyncratic subsets of verbs, before averaging we standardized each child's set of overregularization rates by converting them to z-scores across his or her set of irregular verbs used at least 5 times in the past tense. These aggregate rates are summarized in Table 3.

7.1. Frequency

If overregularization results from a failure to retrieve a listed past tense form, forms with greater memory strength should be more resistant to overregularization. The more often a parent uses a past tense form, the stronger the memory trace for that form should be, and the stronger the association between it and the corresponding stem form. Thus the adult frequency of an irregular verb in its past tense form should be negatively correlated with its overregularization rate for children. Bybee and Slobin (1982a) found a significant negative rank order correlation over verbs between their preschool children's overregularization rates and the frequencies of the verbs in the speech of the preschool children's caretakers. We sought to replicate this effect with our larger sample of children, and to examine it within individual children, to ensure that it is not an averaging artifact.

Three measures of adult frequency were assembled. Frequencies of irregular past tense forms in the speech of the adults talking to Abe are most appropriate for predicting Abe's overregularization rate. For the other children, a larger and more representative measure was obtained by counting forms in the speech of the adults talking to Adam, Sarah, Abe, and Peter. Finally, we used the past tense counts from Francis and Kucera's (1982) corpus of a million words of written text, though these counts are less likely to predict children's behavior because they are from written English addressed to adults. Log frequencies were used in all cases because the frequencies could range over several orders of magnitude (especially for Francis-Kucera figures) and we expected that a frequency difference of 1 versus 10 would have a greater effect than a frequency difference of 1001 versus 1010.

Frequency had a very clear effect. Abe's overregularization rates across verbs correlated significantly with his parents' past tense log frequencies ($r(64) = -.45$, $p < .0001$) and, to a lesser extent (as expected) with Francis & Kucera log frequencies ($r(64) = -.22$, $p < .05$; all p's reported in this section are one-tailed.) The aggregate overregularization rates across 18 children significantly correlate with the aggregate parental frequency counts, $r(34) = -.55$, $p < .0005$, and nonsignificantly correlate with Francis & Kucera frequencies, $r(41) = -.17$, $p > .10$.[7] Of the 18 children, all had overregularization rates across verbs that correlated negatively with the log aggregate parental measure, and 15 of the eighteen had negative correlations with log Francis & Kucera frequency. These sets of individual children's correlation coefficients have means that are significantly less than zero: mean r -.33, $t(17) = -9.56$, $p < .0001$ for log 4-parents frequency); mean r -.15, $t(17) = -4.35$, $p < .0005$ for log Francis-Kucera frequency.

A frequency effect for regularization of irregular verbs is a very pervasive phenomenon. Aside from the effects on overregularization rates in the spontaneous speech of preschoolers documented here and in Bybee and Slobin (1982a), the effect turns up in five other places:

Experimentally elicited production in children: First, Bybee and Slobin (1982a) found that for some subclasses of verbs, frequency correlated negatively with overregularization rates for the experimentally elicited past tense forms in their third grade subjects.

Experimentally elicited speech errors in adults: Bybee and Slobin also documented frequency effects for

[7]Different numbers of verbs figured in the correlations with parental and Francis-Kucera frequency estimates, because no-change verbs and verbs that could be auxiliaries were included in the latter.

some subclasses of verbs in the speech errors of their adult subjects in a time-pressured production experiment.

Historical change: Bybee (1985) examined 33 surviving verbs from three classes of strong verbs in Old English. Fifteen have come through in Modern English as irregular verbs; 18 have become regular. The surviving irregulars have a mean Francis & Kucera frequency of 515 over all their inflectional forms, 137 in the past tense; the regularized verbs have a mean frequency of 21 over all forms, 5 in the past tense.

Coexistence with regularized versions: About 30-40 irregular verbs admit regular past tense forms as more-or-less natural alternatives in casual American speech, for example, *dreamt/dreamed* and *dove/dived* (see Pinker & Prince, 1988, and Ullman and Pinker, 1990, for a list). Stemberger (1989) and Ullman and Pinker (1990) have found that these "doublets" have lower average nonpast stem frequencies than verbs that are exclusively irregular, $t(180) = 2.32, p < .05$.

Preference for the regular form within doublets: Ullman and Pinker (1990) have found that within irregular doublets, the frequency of the irregular past tense form correlates significantly with experimental subjects' relative ratings of the naturalness of the irregular versus the regularized past tense forms.

The pervasiveness of the correlation between frequency and various forms of overregularization strongly suggests that a single psychological phenomenon underlies them all: the strength of the memory trace of the irregular past tense form, and the concomitant probability that the regular process is blocked. Of course this psychological effect does not run through all of these phenomena in identical ways for all verbs. As Bybee and Slobin (1982b) point out, one cannot attribute historical change to children's overregularization errors unless children continue to make the errors through adulthood, which for virtually all irregular verbs in a given generation, they do not. However a weak irregular memory entry in adults can lead to occasional blocking failures, hence regularizations, for the same reason that children overregularize. Presumably this reduces the frequency of the irregular past tense form in the parent generation's speech further, and combined with an overall decline in all tense forms for the verb, it may erode to the point where one generation of children rarely hears it, and hence never cease to overregularize it, at which point it has changed to a regular.

7.2. Phonological Similarity Between Stem and Past

An important fact about irregular verbs is that their past tense forms, though unpredictable in other regards, generally preserve most of the phonological composition of their stems. *Go/went* and *be/was* are exceptions; for the other irregulars such as *come/came, feel/felt,* and *bring/brought*, the past and stem overlap to an extent that would be uncanny if the pair consisted of two arbitrary words linked only as memorized paired associates. Pinker and Prince (1988) point out that any viable theory of irregular morphology must explain this fact, but that the Rumelhart-McClelland model failed to do so.

In some theories of generative grammar (e.g., Chomsky & Halle, 1968; Halle and Mohanon, 1985), the explanation is that irregular pasts are generated by applying to the stem one or more rules that replace a circumscribed substring of phonological segments. It is in the very nature of rules that any segment not changed by the rule is left untouched and hence will automatically appear in the past tense

form; this is how the similarity between the members of an irregular pair is explained.

MacKay (1976) suggested that these rules are applied by speakers on-line when they produce irregular past forms, and that each application consumes a determinate portion of processing resources. When measuring the response times for adult subjects to produce past tense forms when given their stems, he found that verbs with "simple" vowel changes were produced most quickly, followed by regular verbs, followed by verbs with "complex" vowel changes, with verbs with both a vowel change and the *t* suffix being slowest. If MacKay's hypothesis that the psychological complexity of irregulars is predicted by the number of rule applications in the grammatical derivation proposed by irregular rule theories, such effects may also affect children's overregularizations: irregulars with more changes require more rule applications and hence may be harder to produce; when the derivation breaks down, the regular rule steps in.

This hypothesis can be tested by correlating the number of phoneme changes that must be executed to derive the past from the stem with the overregularization rate for that stem. What counts as a "change" will of course depend on one's theory of possible phonological operations, but a reasonable if crude first approximation would be to count each single vowel substitution, consonant substitution, consonant addition, or consonant deletion as one change. For example, in *see-saw*, one phoneme, the vowel, must be replaced; in *seep-swept*, a vowel must be replaced and a consonant suffixed, for a total of two. Hence on average a verb like *sweep* should be overregularized at a higher rate than a verb like *see*. In these calculations, we treated diphthongs as a single phoneme.

We did not find a positive correlation between Number of Phonemes Changed and Overregularization Rate: for Abe, $r(64) = -.19$; for the aggregate of the other children, $r(38) = -.14$.[8] For 13 of the 18 children studied individually, correlations were also negative, as was the mean of the correlations (-.08).

It is possible that there really are no irregular rules, and that the commonalities between stem and past (a heterogeneous set of patterns which range from simple vowel changes to such such severe distortions as *bring/brought*) are to be accounted for by other means. For example, these pairs may have been generated in earlier stages of the language by genuine rules, now defunct; the pairs that were produced by these rules, because they shared phonological material, were easy for learners to memorize, and have preferentially survived over the centuries in Darwinian fashion (see Bybee and Slobin, 1982a, b; Lieber, 1980; Pinker and Prince, 1988).

In particular, Bybee and Slobin point out that phonological overlap between stem and past would help the child link them up as members of a single verb paradigm. They suggest that this is why the class of verbs containing *see/saw, grow/grew, fly/flew, draw/drew,* and *know/knew* were the most prone to overregularization; only a single consonant links the members. If so, it would not be the number of phonemes changed, but the number of phonemes preserved, that should be related to overregularization-proneness. To test Bybee and Slobin's hypothesis, we measured the degree of phonological overlap for all

[8]No-change verbs were included; verbs that were also auxiliaries were excluded.

the irregular pairs by counting the number of shared phonemes. For example, *forget-forgot*, with 5 phonemes preserved, should be overregularized less often than *catch-caught*, with only 1 preserved. Of course every phoneme that is changed is a phoneme that is not preserved, so the two measures correlate negatively, but only -.53, so there should be enough independent variance to test for a negative correlation with overregularization rates even though we just saw that the phonemes-changed measure had no effect. In fact, the correlation is positive for Abe (.32), the aggregate overregularization rate over the other children (.07), the z-transformed mean correlation for these children (.03), and 6 of them individually.

One might note that the irregular with the largest change from stem to past, and the least phonological material preserved, *go/went*, is high in frequency. This suggests that any difficulty in learning more dissimilar past pairs might be compensated for by the higher frequency of such items. In that case the similarity effect might be masked in a simple regression analysis. An effect of phoneme change could be unmasked in a multiple regression analysis using Frequency and Phonemes Changed as predictors; a similar multiple regression could reveal an effect of phoneme preservation. For Abe, Frequency continued to have a significant effect in these analyses ($p < .005$). Both similarity measures showed partial correlation coefficients in the predicted directions, but neither was significant (Phonemes Changed: $F(1,63) = 1.82$, $p > .15$; Phonemes Preserved: $F(1,63) < 1$). Moreover, for both similarity measures the partial correlations are in the nonpredicted direction for the aggregate measure over children and for large numbers of the individual children. In sum, we have failed to find clear evidence that the degree of stem-past similarity or dissimilarity affects the overregularization rate of irregular verbs.

The failure of any measure of stem-past similarity to explain differing overregularization rates is a puzzle. One likely possibility is that the measure of the number of phonemes preserved is so crude as to be misleading an index of psychological similarity. For one thing, the relevant calculation be over distinctive features, not phonemes, so that, for example, the final consonant change of *bend/bent* would count as more similar than that of *make/made*.[9] More importantly, speakers certainly represent words in a format that is more structured than a simple list of segments. In particular, the consonant-vowel skeleton underlying a form is thought to be a distinct level of representation in phonology (see Kaye, 1990), and this would saliently capture the relatedness of *see* and *saw* or *throw* and *threw* even if the sheer number of shared segments was slight. We leave better tests of a possible effect of phonological similarity on overregularization, and more generally, the best explanation of the existence of stem-past similarity, to future research.

7.3. Protection by Families of Similar Irregular Pairs

Not only are irregular verb stems similar to their past tense forms; they tend to be similar to other irregular verb stems that have comparable past tense forms. Irregular verbs fall into clusters such as *sting-stung*, *swing/swung*, *string/string*, and so on. The minor rules posited by some theories of generative morphology (e.g., a rule changing *i* to ^; see Halle & Mohanon, 1985) are meant to explain this second

---

[9]However, we have calculated rough measures using number of distinctive features changed or preserved; these measures do not result in consistent correlations in the predicted direction with overregularization rates.

kind of similarity as well. The form of irregulars is by definition unpredictable on phonological grounds, so the rules must be tagged as applicable only to a fixed list of words, but if the number of rules is smaller than the number of words, the existence of similarity clusters is explained.

However, Pinker & Prince (1988) pointed out problems for such theories. The irregular clusters are held together by far more common features than just the segment changed by the putative rule: *string, sting,* and *swing* share not only an *i*, but a velar nasal as their final consonant and an *s* as part of an initial consonant cluster. Trying to capture these *hypersimilarities* by adding them to the rule as context terms (e.g., "change *i* to ^ in the context *C_ng*") fails in both directions. It falsely includes many forms like *bring/brought* and *sing,* and fails to include verbs that are clearly related to the cluster by family resemblance, such as *stick* (final consonant velar non-nasal) and *spin* (final consonant nasal non-velar); see also Bybee and Slobin (1982a).

In some ways the Rumelhart-McClelland model handles these imperfect partial similarities well: after being trained on 82 irregulars, some of the model's outputs for new irregulars it had not previously encountered were correct, such as *wept, clung,* and *bid,* despite the complex and highly probabilistic nature of the patterns that such generalizations represent. Furthermore the model proved to be highly sensitive to the subregularity that no-change verbs all end in a *t* or *d*, overgeneralizing it to regulars and to other irregulars that end in *t* or *d*. This can be attributed to the fact that the model records the relative frequencies of many different mappings between substrings of stems and substrings of pasts and it superimposes them across the different verbs that exemplified them.

Pinker and Prince (1988), while rejecting Rumelhart and McClelland's suggestion that both regular and irregular forms are generated in a single associative network, conceded that their model might offer insights as to how irregular verbs are stored. If the traditional notion of rote memory for irregular storage is thought of not as an unstructured list of slots, but as some kind of associative network in which recurring similarities are recorded and superimposed, the hypersimilar family resemblance classes can be explained because they contain sets of verbs that are easier to memorize than unrelated singletons, and are prone to occasional generalizations (e.g., *brung, bote*) by analogy. In this interpretation Rumelhart and McClelland would be providing a better model of the irregular rote component of the inflectional system. A related suggestion had been made prior to Rumelhart and McClelland (1986) by Bybee and Slobin (1982a). They suggested that speakers form *schemas* for recognizing typical phonological patterns of irregular past tense forms. Children learn to associate past forms with their stems more easily if they conform to a past tense schema, and they are more likely to select stored forms that conform to a schema when producing past tenses.

If being part of a family of similar irregulars undergoing similar changes strengthens the memory trace of a given irregular form, it should be more resistant to overregularization than more isolated irregulars, holding frequency constant. The prediction that partial regularity blocks overregularization was first suggested by Slobin (1971), and has been further tested by Kuczaj (1977, 1978) and Bybee and Slobin (1982a). The most robust effect is that verbs that end in *t* or *d* are less likely to have *-ed* added, and are more likely to be uttered in no-change form, than verbs without those endings. This is true both for no-change irregulars, leading to improved performance, and for other kinds of irregulars, leading to no-change errors; both kinds are protected from overregularization (see Rumelhart & McClelland, 1986,

and Pinker & Prince, 1988, for reviews.) As mentioned, the Rumelhart-McClelland model duplicated this phenomenon. However, Pinker and Prince (1988) point out that the this effect is potentially so overdetermined that identifying the psychologically active cause or causes is nearly impossible. The no-change class is large, shows an exceptionless hypersimilarity (all verbs ending in *t* or *d*), a single kind of change (none), shares its verb-final consonant with the regulars, and when regularized results in a phonological pattern (adjacent identical stop consonants) that the phonology of English tries to avoid. Thus the existence of an effect of family strength should be tested with other materials.

Bybee and Slobin also showed that children overregularization subclasses with different kinds of vowel changes at different rates. They attributed the differences to different degrees of stem-past similarity, as discussed in the preceding section. But as we have seen, the effects of stem-past similarity are difficult to demonstrate if they exist at all. Furthermore, Pinker and Prince (1988) showed that stem-stem similarity may be the more relevant factor: the overregularization rates for the different vowel change classes correlate well with the number of English irregular verbs sharing the vowel changes that the class members undergo. They suggested that this explains why the Rumelhart-McClelland model model mimicked the ranking of overregularization rates for these subclasses, at least in one stage. However, even here the existence of a stem family effect was not perfectly clear: Bybee and Slobin's subclasses, as interpreted by Rumelhart and McClelland, were heterogeneous and contained many possible contaminants, such as the inclusion of unusual *go-went* in the *blow/grow/know* subclass.

A better test of the family strength effect would eschew the necessarily imperfectly constructed subclasses in favor of a direct measure of the strength of the family members for each irregular verb. While measuring all possible patterns of similarity across all irregulars is not feasible, the following measure should tap an effect of family strength if there is one, because it captures the principle dimension of similarity within families of irregulars (Pinker & Prince, 1988): for each verb we summed the frequency (not the log frequency) of the past tense forms of each of the other irregular verbs whose stems and past tense forms rhyme with those of the verb in question. For example, for *sting-stung* we would add the frequencies of *clung, flung, swung,* and so on. The verb's own frequency was not included; although it surely affects the strength of the family it belongs to, we wanted to see if we could find independent support for a family strength effect, unconfounded by the frequency effect already documented. Because many irregular neighbors like *clung* would be far-fetched candidates for children's lexicons, we actually selected members of irregular word families, and took their frequencies, from the adults' speech in the transcripts of the child in question, using the ten children with individual sets of transcripts. Because the independent variables took on very different values for each child, aggregate measures are not appropriate.

The correlation coefficient between family strength and overregularization rate was -.15 for Abe, and was also negative for 8 of the remaining 9 children; in other words, the higher the frequencies of an irregular verbs' rhymes, the less likely it is to be overregularized. The mean of the ten correlations, -.08, is significantly different from zero, $t(9) = -4.77$, $p < .001$. To ensure that this effect cannot be attributed to a confound with the frequency of each verb, we held frequency constant in a partial correlation analysis; the mean partial correlation coefficient stayed at -.08 and remained negative for 9 of the 10 children. We conclude that there is a reliable, though weak, effect whereby verbs are protected from overregularization to the extend that they are phonologically similar to other verbs (weighted by their frequencies)

displaying the same irregular pattern.

7.4. Attraction to Families of Similar Regular Verbs

The preceding analysis confirms the hypothesis of Slobin, Bybee, and Kuczaj that partial regularity blocks overregularization; it is consistent both with the Rumelhart-McClelland model and with Pinker and Prince's augmentation of the traditional rote-rule model in which the rote component has associative-memory-like properties. A test that distinguishes the latter two models is whether families of similar *regular* verbs pull an irregular *toward* overregularization, in the same way that families of similar irregulars pull it away. Since in the traditional rote & rule model regular past tense forms are not stored but generated by a rule, under the simplest hypothesis there is no way that regular past tense forms could attract irregulars. That this theoretical difference does in fact distinguish rule-based theories from the Rumelhart-McClelland model was shown by the behavior of the model on newly presented regular verbs. At asymptote the model erred on 33% of the regular verbs it was tested on, producing no output at all for six that were dissimilar from those in its training set such as *jump* and *pump*. In contrast, adults easily regularize highly unusual sounding nonce forms such as *ploamph* or *keelth* (Prasada and Pinker, 1990), and the 2-5 year old Abe left no regular verb in past tense contexts unmarked, including his own unusual inventions *eat lunched, bonked, borned*, and *axed* (Kuczaj, 1977).

To test whether families of regular verbs pull similar irregulars toward overregularization, we first extracted the 1826 regular verbs rhyming with the irregulars that were listed in an on-line version of Webster's 7th Dictionary, which contains phonological representations of 8217 verbs. The sums of the frequencies of the regulars in a parent's speech that rhymed with each of the child's irregulars were tallied. If families of regular verbs pull similar irregulars toward overregularization, correlations between regular family strength and overregularization rate should be positive. Instead, the correlations were negative for 6 of the children (including Abe), and positive for the other 4; the sample of normalized correlations did not differ significantly from zero (mean r = .04, $t(9) < 1$). Because a putative regular family might span a large set of roughly-similar verbs, rather than a smaller set of highly-similar verbs, we also calculated a more inclusive family strength measure, in which we added to each irregular's family of similar regulars the regular verbs from the subsample of 1826 that merely shared a final consonant with the irregular. Here again, regular cluster strength did not correlate positively with overregularization: the correlations were negative for 4 of the 10 children, including Abe, and though the mean of the correlations following normalization was .05, this mean did not differ significantly from zero, $t(9) = 1.09, p > .25$.

We conclude that there is no evidence that irregular verbs are drawn toward overregularization by families of similar regulars, though they seem to be protected against overregularization by families of similar irregulars. This supports Rumelhart and McClelland's assumption that irregular patterns are stored in an associative memory, but fails to support their assumption that the regular pattern is stored in the same system.

8. Summary and Conclusions

The facts of overregularization can be summed up simply. Overregularization occurs at a constant low rate, affecting all irregular verbs, during the preschool and early school years. Its overall rate is independent of changes in the mixture of regular and irregular verbs in the child's speech, the child's parents' speech, or the child's vocabulary. Instead it is related to the degree of obligatoriness in the child's marking of tense in general: when irregular verbs are reliably marked for tense, overregularization errors appear. Not all verbs are overregularized equally often; the chief determinant of overregularization-proneness is the verb's frequency in parental speech; high frequency verbs are overregularized less often. There is a small effect of the strength of the verb's phonological neighborhood; clusters of similar irregular verbs protect one another from overregularization. In contrast, clusters of similar regular verbs do not pull an irregular toward overregularization.

These facts can be accounted for by a simple theory: the child stores irregular past tense forms in a rote memory system, in which the strength of a memory trace is monotonically related to the frequency with which it is encountered. In addition, this memory system has some of the properties of an associative network: stems-past pairs displaying similar relationships reinforce each other.[10] Regular past tense forms, in contrast, are generated by a mental concatenation operation that affixes a suffix to a stem. Because this rule can always be applied on-line, regularly inflected forms need never be stored; because it simply adds an affix to the end of a stem with unspecified properties, the similarity of a given stem to previously encountered ones plays no role.[11] The two systems interact in a simple way: the retrieval of a stored irregular blocks the application of the regular rule.

The fact that overregularizations are a small minority of irregular past tense utterances at all stages shows that this principle is active in the child as soon as there is evidence for two modes of inflection at all. When overregularizations do occur, they are straightforwardly explained as a failure to retrieve the irregular past form (or, for past+*ed* errors, its "past" feature) in real time. This tendency is related to the frequency of the form in an obvious way. In the extreme case, an irregular form that has been attended to with zero frequency (e.g., *shend* for adults, and many irregular verbs for young children) will have no memory trace and hence will be retrieved from memory with zero probability, and will always be overregularized if the form is to be tense-marked in some way. An irregular form that has been heard once has a weak memory trace and hence a probability of being retrieved that is greater than zero but less than one. Irregulars that have been heard more times have correspondingly stronger memory traces and lesser overregularization probabilities; irregulars that have been heard thousands of times will be successfully retrieved virtually always. The learnability problem of recovering from errors is solved by a Blocking principle that operates throughout development, fed by irregular forms whose potency increases with increasing exposures during development.

Thus the explanation of the phenomenon of overregularization has three parts: imperfect memory

[10]This same property leads with low but nonzero probability to irregular generalizations such as *brang* and *thunk*.

[11]The variation in the phonetic form of the regular affix, whereby *d* surfaces as *t* following unvoiced consonants and *id* following coronal stops, are attributable to general phonological processes operating throughout English, not to the regular process itself; see Pinker and Prince, 1988).

retrieval, obligatory marking of tense, and possession of a regular process with universal applicability. Though the first postulate reflects a possibly uninteresting memory limitation of the child, the other two reflect quite remarkable linguistic accomplishments.

As Slobin (1982) and Pinker (1982, 1984) point out, obligatory grammatical constraints pose difficult learning problems for the child. The fact that an inflection is obligatory means that there are no pragmatic cues to the semantic features that the inflection is encoding; parents must encode the pastness of a past event, regardless of how relevant it is in the conversational context. Moreover once the child somehow has figured out that past tense inflection encodes past tense, if he or she mistakenly assumes that it is optional, no parental input short of negative evidence can contradict the assumption. Slobin and Pinker thus suggest that the child is innately prepared to consider obligatory tense-marking as a possible constraint in the language to be acquired; such a hypothesis is easily disconfirmed in systems where inflection is in fact optional. Our seemingly homely explanation for overregularization in terms of retrieval failure depends on the child having solved these daunting learnability problems (see Pinker, 1984, for an explicit hypothesis as to how the child solves them; the problems are of course finessed in network simulations which are fed correct stem-past pairs in isolation.) If the three-year old child did not consider tense marking to be obligatory, the failure to retrieve an irregular past in a past tense context would not automatically lead to an overregularization; the child could simply leave the verb unmarked.

Moreover, the *intention* to mark a stem for past tense is not sufficient to explain *success* at doing so. The second noteworthy linguistic achievement of the late two-year-old is possession of a process that is capable of yielding an inflected output form for any verb, no matter how strongly linked with an idiosyncratic irregular form, and regardless of whether a family of similar regular forms is available to serve as an analogy-supporting model. As Pinker and Prince (1988) point out, a rule that simply concatenates an affix with a stem, characterized in terms of a variable standing for any stem, rather than particular patterns of the possible phonological contents of stems, easily provides this capability. A set of associations between stem phonology and past tense phonology is inherently tied to the patterns it has been trained on. While such models could, in principle, approximate the unlimited applicability of a rule by training it on a set of regular stems that span enough of the phonological space of English to cover all cases, in practice this ability is compromised by the necessity of curbing links to the regular ending in order to avoid application to the irregulars, and thus the Rumelhart-McClelland model failed to display the appropriate generalization abilities for novel regulars. The children we have studied (most notably Abe), in managing to come out with a past tense form almost 100% of the time when called for, despite less than perfect memory retrieval, clearly have mastered an inflectional process of very wide applicability. If this regular process was not capable of applying to arbitrary irregular stems (and in fact 89% of Abe's irregulars were overregularized at least once) or if it depended on the existence of similar high frequency regulars (which in fact it did not), when faced with irregular retrieval failure Abe would be left with no choice but to utter the stem, even if his grammar called for a past tense form.

Of course there are aspects of overregularization that remain to be explained. In particular, there is considerable unexplained variance in exactly which verbs are overregularized at which rates and ages. These phenomena may involve aspects of memory microdynamics that are poorly understood at present, and the data from currently available speech samples are too scanty to shed much light on them.

We have shown that existing connectionist models restricted to a single associative network cannot account for the facts concerning children's level, onset, and lexical distribution of overregularization. However we are not not claiming that no connectionist model is capable of doing so. The set of possible connectionist models encompasses a wide variety of relative propensities for rote memory versus generalization, and includes models in which the balance between these tendencies changes during a training run. These tendencies are influenced by a variety of design parameters left open to the network creator, such as the coding scheme for the input features, their degree of probabilistic blurring, the topology of the network (e.g., the number of hidden layers, and the number of nodes in each one), the learning rate, the momentum factor, the temperature, the training schedule, and others. Such properties are not mandated by any general connectionist theory but are implemented as innate features of a given model by the designer, usually after a period of trial and error to see what works (see, e.g., Plunkett and Marchman, 1990; MacWhinney, 1990). Thus someone might find design parameters that allow some connectionist model to simulate children by displaying the right rote and regularization modes at different points in development using a unitary network. But in such a case the explanation of why children overregularize is not that they are connectionist networks, but that the mechanism they possess, connectionist or otherwise, is designed to display the distinct rote and rule behavior patterns such as those we have documented here.[12]

Moreover there is one set of facts that favors a dual-mechanism (rule & rote) model over any unitary network, no matter how well it is designed to display rulelike and rotelike modes. Linguistic research has shown that regular and irregular inflected forms differ *qualitatively*, in terms of their sensitivity to qualitative grammatical distinctions and their relation to other grammatical processes. In at least three cases children have been shown to respect these qualitative grammatical distinctions in their patterns of regularization and overregularization.

(1) Gordon (1986) noted that English compounds can contain irregular plurals (e.g., *mice-infested*) but not regular plurals (e.g., **rats-infested*), a consequence of the fact that such compounds are composed of stems stored in the lexicon, not complex forms created by an inflectional rule. In an elicitation experiment, he found that children produced novel compounds containing irregular plurals (*mice-eater*) but never containing regular plurals (**rats-eater*), nor did they form compounds out of their own overregularizations – children who said *mouses* would nonetheless avoid saying **mouses-eater*.

(2) Stromswold (1990) notes that English has three irregular verbs that are morphologically identical to auxiliaries: *do* (as in *do a favor*), possessional *have*, and copula *be*. Each of these verbs is identical to an auxiliary not only in the stem form but in every single one of its irregularly inflected versions: *do*, whether used as a main verb or as the auxiliary in negations, inversions, and emphatics, has the irregular forms *does, did*; perfect *have*, like possessional *have*, maps onto the irregular forms *has* and *had*, and *be*, whether serving as a copula verb, a progressive auxiliary, or a passive auxiliary, has the irregular forms *am, is, are, was, were*. Clearly there are too many of these parallelisms to be coincidental and one can

---

[12] For example, although MacWhinney (1990) presents a connectionist model of past tense acquisition, he notes that his explanation for children's U-shaped development has nothing to do with that network, but is similar to the explanation advocated in MacWhinney (1978), Pinker and Prince (1988), and this paper: that children memorize individual lexical entries using a mechanism separate from the one that generates rule-governed alternations.

conclude that the irregular forms of the main verb and of the auxiliary versions are stored in the same mechanism. But their susceptibility to overregularization is qualitatively different: in a sample of 40,000 child sentences containing these verbs, Stromswold found that the main verb versions are overregularized at rates comparable to those we have found here, whereas the auxiliary versions of the same verbs were *never* overregularized. This extreme conservatism was predicted, on the grounds of learnability considerations and other developmental data, by the theory in Pinker (1984) according to which auxiliary verbs are marked with a feature that renders them untouchable by regular morphological operations. The main point is that children are capable of treating two verb forms identically in terms of irregular inflectional patterns, but qualitatively differently in terms of regular inflectional patterns.

(3) Kim, Pinker, Prince and Prasada (in press) review evidence from linguistic research showing that verbs derived from nouns and adjectives are regular, even if homophonous to an irregular verb. For example, *to high-stick* (= "hit with a stick held high") has the past tense *high-sticked*, not *high-stuck*. This fact, demonstrated to be extremely robust in naive subjects faced with inflecting novel denominal verbs, is a consequence of the fact that possessing an irregular past tense form is a property that holds of verb *roots*, not verbs in general (unless they have an irregular verb root as their head); verbs formed from nouns have a noun as their head, not an irregular verb root, and hence reliably regularize. So in order to predict a verb's past tense form, its phonological properties do not suffice; its morphological structure must be input to the relevant mechanisms. There is some evidence that this is true of children as well: Kim, Marcus, Hollander, and Pinker (in preparation) have shown that children regularize denominal verbs homophonous with irregulars, such as *to fly* = "to cover a piece of paper with flies") far more often than they regularize the irregular verb itself.

Thus it is necessary to attribute children's overregularizations to a different mechanism than their irregulars, on the basis of the qualitatively different inputs and outputs the two patterns implicate: the production of irregulars but not overregularizations feeds compound formation; the production of irregulars but not overregularizations is fed by auxiliary verbs; the production of regularizations but not irregulars is fed by verbs derived from nouns. We have shown that exactly this kind of distinction among memory and rule mechanisms provides an extremely simple account for a huge variety of facts about the rate, onset, and lexical properties of overregularization.

9. References

Anisfeld, M. & Gordon, M. (1968) On the psychophonological structure of English inflectional rules. *Journal of Verbal Learning and Verbal Behavior, 7*, 973-979.

Anisfeld, M. & Tucker, G. R. (1967) The English pluralization rules of six-year old children. *Child Development, 38*, 1201-1217.

Anisfeld, M., Barlow, J., & Frail, C. M. (1968) Distinctive features in the pluralization rules of English speakers. *Language and speech, 11*, 31-37.

Aronoff, M. (1976) *Word formation in generative grammar*. Cambridge, Mass.: MIT Press.

Bateman, W. G. (1916) The language status of three children at the same ages. *Pedagogical Seminary, 23*, 211-240.

Berko, J. (1958) The child's learning of English morphology. *Word, 14*, 150-177.

Bever, T. G. (1975) Psychologically real grammar emerges because of its role in language acquisition. In D. Dato (Ed.), *Developmental psycholinguistics: Theory and applications. Georgetown University Round Table on Languages and Linguistics.* Washington, DC: Georgetown University Press.

Bloom, L. (1973) *One word at a time: The use of single word utterances before syntax.* The Hague: Mouton.

Bohannon, J. N. & Marquis, A. L. (1977) Children's control of adult speech. *Child Development, 48*, 1002-1008.

Bowerman, M. (1982) Starting to talk worse: clues to language acquisition from children's late errors. In S. Strauss, ed., *U-shaped behavioral growth*. New York: Academic Press.

Bowerman, M. (1987) Commentary: Mechanisms of language acquisition. In B. MacWhinney, ed., *Mechanisms of language acquisition*. Hillsdale, N.J.: Erlbaum.

Brown, R. & Bellugi, U. (1964) Three processes in the child's acquisition of syntax. In E. Lenneberg (Ed.), *New directions in the study of language.* Cambridge, MA: MIT Press.

Brown, R., and Hanlon, C. (1970) Derivational complexity and order of acquisition in child speech. In J. R. Hayes, ed., *Cognition and the development of language.* New York: Wiley.

Brown, R. (1973) *A first language: The early stages.* Cambridge, Mass.: Harvard University Press.

Bryant, B. & Anisfeld, M. (1969) Feedback versus no-feedback in testing children's knowledge of English pluralization rules. *Journal of Experimental Child Psychology, 8*, 250-255.

Burnham, K. P. and Overton, W. S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika, 65*, 625-33.

Burnham, K. P. and Overton, W. S. (1979) Robust estimation of population size when capture probabilities vary among animals. *Ecology 60*, 927-936.

Bybee, J. L. & Slobin, D. I. (1982a) Rules and schemes in the development and use of the English past tense. *Language, 58*, 265-289.

Bybee, J. L. & Slobin, D. I. (1982b) Why small children cannot change language on their own: Suggestions from the English past tense. In A. Ahlqvist (Ed.), *Papers from the 5th Internation Conference on Historical Linguistics. (Current Issues in Linguistic Theory Vol. 21, Amsterdam Studies in the theory and history of linguistic science IV.)* Philadelphia/Amsterdam: John Benjamins.

Bybee, J. L. (1985) *Morphology: a study of the relation between meaning and form.* Philadelphia: Benjamins.

Carlton, L. E. (1947) Anomalous preterite and past participle forms in the oral language of average fourth grade children. *American Speech, 22*, 40-45.

Carroll, J. B. (1961) Language development in children. In S. Saporta (Ed.), *Psycholinguistics: A book of readings.* New York: Holt Rinehart & Winston.

Cazden, C. B. (1968) The acquisition of noun and verb inflections. *Child Development, 39*, 433-448.

Cazden, C. B. (1972) *Child language and education.* New York: Holt Rinehart & Winston.

Chamberlain, A. F. (1906) Preterite forms, etc., in the language of English-speaking children. *Modern Language Notes, 21*, 42-44.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal capturability. *Biometrics, 43*, 783-791.

Chomsky, N. (1959) A Review of B. F. Skinner's "Verbal Behavior". *Language*, 3, 26-58.

Clark, E. V. (1982) The young word maker: a case study of innovation in the child's lexicon. In E. Wanner and L. R. Gleitman, eds., *Language acquisition: The state of the art.* New York: Cambridge University Press.

Clark, E. V. (1987) The Principle of Contrast: A constraint on language acquisition. In B. MacWhinney, ed., *Mechanisms of language acquisition.* Hillsdale, N.J.: Erlbaum.

Cox, M. V. (1989) Children's over-regularization of nouns and verbs. *Journal of Child Language, 16*, 203-206.

Derwing, B. & Baker, W. (1979) Recent research on the acquisition of English morphology. In P. Fletcher & M. Garman (Eds.), *Language Acquisition.* New York: Cambridge University Press.

Edwards, M. L. (1970) An annotated bibliography on the acquisition of English verbal morphology. *Ohio State University Working Papers in Linguistics, 4*, 149-164.

Ervin, S. M, & Miller, W. R. (1963) Language development. In H. W. Stevenson (Ed.), *Child psychology: The Sixty-Second Yearbook of the National Society for the Study of Education, Part I.* Chicago: University of Chicago Press.

Ervin, S. M. (1964) Imitation and structural change in children's language. In E. H. Lenneberg (Ed.), *New directions in the study of language.* Cambridge: MIT Press.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985) Relations among regular and irregular morphologically related words in the lexicon as related by repetition priming. *Memory and Cognition, 13,* 241-255.

Francis, N. & Kucera, H. (1982) *Frequency analysis of English usage: Lexicon and grammar.* Boston: Houghton Mifflin.

Gleason J. B. (1980) The acquisition of social speech and politeness formulae. In H. A. R. Giles (Ed.) *Language: Social psychological perspectives.* Oxford: Pergammon.

Gold, E. M. (1967) Language identification in the limit. *Information and Control, 16,* 447-474.

Gordon, P. (1986). Level-ordering in lexical development. *Cognition, 21,* 73-93.

Gordon, P. (1990) Learnability and feedback. *Developmental Psychology 26,* 217-220.

Grimshaw, J., and Pinker, S. (in press) Positive and negative evidence in language acquisition. (Commentary on D. Lightfoot's "The child's trigger experience: 'Degree-0' learnability.") *Behavioral and Brain Sciences, 12,* 24.

Grimshaw, J. & Rosen, S. T. (1990) Knowledge and obedience: The developmental status of the binding theory. *Linguistic Inquiry, 21,* 187-222.

Guillaume, P. (1927) Le developpement des elements formels dans le langage de l'enfant. *Journal de Psychologie, 24,* 203-229. Translated as "The development of formal elements in the child's speech." In C. A. Ferguson & D. I. Slobin (1983) (Eds), *Studies of child language development.* New York: Holt Rinehart & Winston.

Hall, W. S. & Tirre, W. C. (1979) *The communicative environment of young children: Social class, ethnic, and situational differences.* Urbana: University of Illinois Center for the Study of Reading.

Halle, M. & Mohanon, K. P. (1985) Segmental phonology of modern English. *Linguistic Inquiry, 16,* 57-116.

Higginson (1985) Fixing-assimilation in language acquisition.

Horn, M. D. (1927) The thousand and three words most frequently used by kindergarten children. *Childhood Education, 3,* 118-122.

Ingram., D. (1989) First language acquisition: Method, description, and explanation. New York: Cambridge University Press.

Kaye, J. (1990) *Phonology: A cognitive view.* Hillsdale, NJ: Erlbaum Associates.

Kim, J. J., Marcus, G. F, Hollander, M., & Pinker, S. (in prep.) On children's sensitivity to morphological derivation of words.

Kim, J. J., Pinker, S., Prince, A. Prasada, S. (In press) Why no mere mortal has ever flown out to center field. *Cognitive Science.*

Kiparsky, P. (1982). Lexical phonology and morphology. In I. S. Yang (ed.), *Linguistics in the morning calm.* Seoul: Hansin, 3-91.

Kucera, H. and N. Francis (1967) *Computational analysis of present-day American English.* Providence: Brown University Press.

Kuczaj, S. A. (1976) *-ing, -s, & -ed*: A study of the acquisition of certain verb inflections. Doctoral dissertation, Department of Psychology, University of Minnesota.

Kuczaj, S. A. (1977) The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior, 16,* 589-600.

Kuczaj, S. A. (1978) Children's judgments of grammatical and ungrammatical irregular past tense verbs. *Child Development, 49,* 319-326.

Kuczaj, S. A. (1981) More on children's initial failure to relate specific acquisitions. *Journal of Child Language, 8,* 485-487.

Lachter, J. & Bever, T. G. (1987) The relation between linguistic structure and associative theories of language learning -- A constructive critique of some connectionist learning models. *Cognition, 28,* 195-247.

Lenneberg, E. H. (1964) A biological perspective on language. In E. H. Lenneberg (Ed.), *New directions in the study of language.* Cambridge: MIT Press.

Lieber, R. (1980) On the organization of the lexicon. Doctoral dissertation, Department of Linguistics and Philosophy, MIT.

Lorge, I. & Chall, J. (1963) Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *The Journal of Experimental Education, 32,* 147-157.

MacKay, D. (1976) On the retrieval and lexical structure of verbs. *Journal of Verbal Learning and Verbal Behavior, 15,* 169-182.

MacWhinney, B. & Snow, C. E. (1985) The Child Language Data Exchange System. *Journal of Child Language, 12,* 271-296.

MacWhinney, B. (1978) Processing a first language: the acquisition of morphophonology. *Monographs of*

the Society for Research in Child Development, 43.

MacWhinney, B. (1990) Implementations are not conceptualizations: Revising the verb learning model. Unpublished manuscript, Department of Psychology, Carnegie-Mellon University.

MacWhinney, B. & Snow, C. (1990) The Child Language Data Exchange System: an update. *Journal of Child Language, 17*, 457-472.

Maratsos, M. (1987) Gift of tongues. *Times Literary Supplement* January 2, 19.

Marchman, V. (1988) Rules and regularities in the acquisition of the English past tense. *Center for Research on Language Newsletter, 2(4)*, University of California, San Diego.

McNeill, D. (1966) Developmental psycholinguistics. In F. Smith and G. Miller, eds., *The Genesis of Language*. Cambridge, Mass.: MIT Press.

Menyuk, P. (1963) A preliminary evaluation of grammatical capacity in children. *Journal of Verbal Learning and Verbal Behavior, 2*, 429-439.

Miller W. R. & Ervin, S. M. (1964) The development of grammar in child language. In U. Bellugi & R. Brown (Eds.), *The acquisition of language. Monographs of the Society for Research in Child Development, 29*, 9-39,

Miller, G. A. (1977) *Spontaneous apprentices -- children and language.* New York, N.Y.: Seabury Press.

Moe, Hopkins, & Rush (1982). The Vocabulary of First-Grade Children. Charles C. Thomas Publisher. Springfield, Ill.

Morgan, J. L., and Travis, L. L. (1989) Limits on negative information on language learning. *Journal of Child Language, 16*, 531-552.

Osherson, D. N., Stob, M., and Weinstein, S. (1985) *Systems that learn.* Cambridge, Mass.: MIT Press.

Otis, D. L., Burnham K. P., White G. C. and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlife Monographs, 63*, 1-135.

Pinker, S. (1979) Formal models of language learning. *Cognition, 7*, 217-283.

Pinker, S. (1982) A theory of the acquisition of lexical interpretive grammars. In J. Bresnan, ed., *The mental representation of grammatical relations.* Cambridge, Mass.: MIT Press.

Pinker, S. (1984) *Language learnability and language development.* Cambridge, MA: Harvard University Press.

Pinker, S., and Prince, A. (1988) On language and connectionism: Analysis of a Parallel Distributed Processing model of language acquisition. *Cognition* 28: 73-193.

Pinker, S., Lebeaux, D. S., and Frost, L. A. (1987) Productivity and constraints in the acquisition of the passive. *Cognition* 26: 195-267.

Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure.* Cambridge, MA: MIT Press.

Plunkett, K., & Marchman, V. (in press) U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition.*

Plunkett, K. & Marchman, V. (1990) From rote learning to system building. Technical Report 9020, Center for Research in Language, University of California, San Diego.

Prasada, S. & Pinker, S. (1990) Similarity-based and rule-based generalizations in inflectional morphology. Unpublished manuscript, Department of Brain and Cognitive Sciences, MIT.

Prasada, S. Pinker, S, and Snyder, W. (1990) Some evidence that irregular forms are retrieved from memory but regular forms are rule-generated. Paper presented to the Annual Meeting of the Psychonomic Society, New Orleans, November 23-25.

Reich, P. A. (1986) *Language development.* Englewood Cliffs, NJ: Prentice-Hall.

Rumelhart, D. E. & McClelland, J. L. (1986b) On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group, *Parallel distributed processing: Explorations in the microstructure of cognition. Volume 2: Psychological and biological models.* Cambridge, MA: Bradford Books/MIT Press.

Rumelhart, D. E. & McClelland, J. L. (1987) Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Sachs, J. (1983) Talking about there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson (Ed.) *Children's Language, Vol 4.* Hillsdale, NJ: Erlbaum.

Sampson, G. (1987a) A turning point in linguistics. *Times Literary Supplement,* June 12, 643.

Sampson, G. (1987b) Parallel distributed processing. *Times Literary Supplement,* October 15.

Seashore, R. H. & Eckerson, L. D. (1940) The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology, 31,* 14-38.

Seber, G. A. F. (1986). A review of estimating animal abundance. *Biometrics, 42,* 267-292.

Slamecka, N. J. (1969) Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology, 81,* 557-560.

Slobin, D. I. (1971) On the learning of morphological rules: A reply to Palermo and Eberhart. In

D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium*. New York: Academic Press.

Slobin. D. I. (1982) Universal and particular in the acquisition of language. In E. Wanner and L. R. Gleitman, ed., *Language Acquisition: the State of the Art*. Cambridge: Cambridge University Press.

Slobin, D. (1978) A case study of early language awareness. In A. Sinclair, R. J. Jarvella, and W. J. M. Levelt (Eds.), *The child's conception of language*. New York: Springer-Verlag.

Slobin, D. I. (1973) Cognitive prerequisites for the development of grammar. In C. Ferguson and D. I Slobin, ed., *Studies of Child Language Development*. New York: Holt, Rinehart and Winston.

Smith, M. E. (1933) Grammatical errors in the speech of preschool children. *Child Development, 4*, 183-190.

Smith, M. E. (1935) A study of some factors influencing the development of the sentence in preschool children.

Smolensky, P. (1988) The proper treatment of connectionism. *Behavioral and Brain Sciences*.

Stanners, R. F., Neiser, J. J., Hernon, W., P., & Hall, R. (1979) Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior, 18*, 399-41.

Stemberger, J. P. (1989) The acquisition of morphology: analysis of a symbolic model of language acquisition. Unpublished manuscript, Department of Linguistics, University of Minnesota.

Stemberger, J. P. (1982) Syntactic errors in speech. *Journal of Psycholinguistic Research* 11: 313-345.

Stemberger, J. P. & MacWhinney, B. (1986) Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition, 14*, 17-26.

Strauss, S. (Ed.) (1982) *U-shaped behavioral growth*. New York: Academic Press.

Stromswold, K. J. (1990) Learnability and the acquisition of auxiliaries. Doctoral Dissertation, Department of Brain and Cognitive Sciences, MIT.

Talmy, L. (1985) Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen, ed., *Language typology and syntactic description*. Vol. III: *Grammatical categories and the lexicon*. New York: Cambridge University Press.

Templin, M. (1957) *Certain language skills in children*. Minneapolis: University of Minnesota Press.

Ullman, M. & Pinker, S. (1990) Why do some verbs not have a single past tense form? Paper presented to the Fifteenth Annual Boston University Conference on Language Development, October 19-21.

Warren-Leubecker A. (1982) Sex differences in speech to children. Master's thesis, Georgia Institute of Technology.

Wexler, K., and Culicover, P. (1980) *Formal principles of language acquisition.* Cambridge, MA: MIT Press.